sustainability

MDPI

Article

Assessing the Impact of China's Industry on Health: A Dynamic Three-Stage Bootstrap Model with an Endogenous Fuzzy Directional Distance Function

Qinghua Pang [1], Xiaopeng Wei [2,*], Lina Zhang [1] and Yanjie Shang [3]

[1] School of Economics and Finance, Hohai University, Changzhou 213022, China; pangqh@hhu.edu.cn (Q.P.); 20191001@hhu.edu.cn (L.Z.)
[2] School of Mathematics, Hohai University, Nanjing 211100, China
[3] School of Management, Jiangsu University of Technology, Changzhou 213022, China; jsczsyj123@163.com
* Correspondence: weixiaopeng1124@163.com; Tel.: +86-17887803015

Abstract: Evaluating the industrial eco-system is essential for promoting resource efficiency, environmental development, and public health protection. However, traditional Data Envelopment Analysis (DEA) methods are prone to biases and rely on precise numerical values, which pose significant challenges when evaluating the industrial ecosystem in China, particularly given the prevalent data uncertainties and environmental complexities. Additionally, DEA models are typically static, fail to capture the long-term trends and dynamic characteristics of the industrial ecosystem. To address these issues, this study proposes a dynamic three-stage model with an endogenous fuzzy directional distance function. The model integrates the generalized smooth bootstrap method and fuzzy comparison techniques to correct errors and account for data uncertainties, improving the accuracy and scientific validity of eco-efficiency assessments. Chinese enterprises face challenges such as excessive resource consumption, environmental pollution, and health risks, necessitating a more comprehensive and flexible evaluation system to adapt to the complex and dynamic nature of the industrial ecosystem. By focusing on industrial production, environmental governance, and health threats (IPEGHT) in China, the research aims to provide a robust framework for enhancing industrial eco-efficiency in a dynamic and uncertain environment. The results show: (1) The proposed method for selecting endogenous direction vectors is applied to the industrial production stage, thereby identifying the subsequent improvement directions for industrial production in various provinces. (2) The idea of integration is adopted to introduce the generalized smooth bootstrap method and fuzzy mathematics, to conduct a precise and comprehensive evaluation of the IPEGHT industrial eco-system. (3) From 2011 to 2021, China's IPEGHT industrial eco-system efficiency showed a step-like distribution, and attention needs to be paid to environmental governance. The proposed method' application can facilitate the integration of DEA and fuzzy mathematics, and fosters a more sustainable industrial eco-system.

Keywords: three-stage fuzzy DEA; bootstrap; endogenous directional vector; industrial production; environment government; health threat

check for updates

Academic Editor: Hirofumi Fukuyama

Received: 28 November 2024
Revised: 22 December 2024
Accepted: 30 December 2024
Published: 14 January 2025

Citation: Pang, Q.; Wei, X.; Zhang, L.; Shang, Y. Assessing the Impact of China's Industry on Health: A Dynamic Three-Stage Bootstrap Model with an Endogenous Fuzzy Directional Distance Function. *Sustainability* **2025**, *17*, 608. https://doi.org/10.3390/su17020608

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The acceleration of industrialization has increasingly underscored environmental pollution and health issues, posing significant challenges to sustainable development. The traditional industrial model, characterized by high energy consumption and emissions,

has imposed great pressure on natural resources while exacerbating air, water, and soil pollution, which posing grave threats to human health (Yang et al., 2024) [1]. According to the World Health Organization (WHO, 2023), approximately 9 million people worldwide died annually from environmental pollution, with China alone recording approximately 140 deaths per 100,000 people in 2022 attributed to environmental or household air pollution. Furthermore, the World Bank (2023) estimated that air and water pollution accounted for a 4.8% loss in China's gross domestic product. The United Nations' sustainable development goals (SDGs), especially SDG 3, 6, 7, and 12, intersect in key areas: industrial production, environmental governance, and health threats. These goals emphasize the need for sustainable practices that safeguard human health and well-being, and promote environmental resilience and equitable access to essential resources and services.

As one of the main energy consumers, China still faces an astonishing overall scale of resource use due to its large population base, exerting unprecedented pressure on the natural environment. The high-intensity resource utilization model not only has exacerbated environmental burdens, but also presented China with more intricate and significant challenges in balancing industrial production, environmental governance, and health protection. Importantly, the large population size also means that China bears a significant responsibility in ensuring the health of its people. According to data from the National Bureau of Statistics released in 2021, from 2011 to 2021, China's wastewater discharge and solid waste generation have increased by 118.7% and 62.6%, respectively, while emissions of air pollutants have severely exceeded standards, nearing 18 million tons annually, highlighting the severity of the environmental situation.

China is actively promoting economic transformation and upgrading. Evaluating the industrial eco-efficiency of various provinces can optimize resource allocation, motivate enterprises to enhance environmental governance, and propel China's industrial sector towards a more efficient, environmentally friendly, and sustainable path. Furthermore, in the context of escalating global environmental awareness, improving industrial eco-efficiency not only satisfies the international market's demand for green products, enhancing the international competitiveness of Chinese products, but also effectively tackles environmental pollution and health threats, thereby enhancing the quality of life and well-being of the populace. Against this backdrop, the traditional mindset of prioritizing economic growth is gradually being abandoned, and China is placing greater emphasis on the integrated development of the economy, environment, and society. Industrial production drives economic development, while it also generates significant amounts of wastewater, exhaust gases, and solid waste, which urgently need to be properly managed. Unprocessed industrial emissions pose direct and indirect threats to human health. From a sustainable development perspective, industrial production (IP), environmental governance (EG), and health threats (HT) can be considered an interconnected industrial eco-system (IPEGHT), which addresses key issues such as SDG 3, 6, 7, and 12. Therefore, it is crucial to conduct an appropriate evaluation of the efficiency of the IPEGHT eco-system, as it will not only reveal the current state of China's sustainable development efforts, but also provide invaluable insights and strong support for future development strategies.

Data Envelopment Analysis (DEA) is a non-parametric method used for evaluating the efficiency of multi-input and multi-output systems. The DEA method has been widely adopted in the realm of ecological efficiency studies, and network DEA methods delve even deeper into the complex internal structures of decision-making units (DMUs). Tone & Tsutsui (2014) [2] not only explored issues of intertemporal carry-over in network DEA models, but also considered the dynamic efficiency of DMUs. The Directional distance function (DDF) is an extension of DEA that allowing non-radial improvements and provides more flexible efficiency evaluations. For example, in environmental efficiency assessments,

DDF can simultaneously optimize pollution reduction and output increase. The use of an endogenous directional vector allows the DDF model to flexibly adjust the evaluation direction based on the actual input and output data of each DMU, accurately reflecting its relative efficiency. Traditional DEA models typically employ exogenous directional vectors, which often requires researchers to set them based on experience or theoretical assumptions, potentially bringing subjective bias (Song & Wang, 2018) [3]. The endogenous directional vector, driven by data, can reduce the influence of subjective judgment, enhancing the objectivity and reliability of evaluation results. Through the endogenous directional vector, the DDF model can more precisely pinpoint potential directions and areas for efficiency improvement for each DMU, which is of significance for formulating targeted improvement strategies and policies. However, the selection of endogenous directional vectors in existing literature is can often complex. In practical applications, the DEA method faces a series of challenges, particularly the issue of sample selection bias. When dealing with small sample data, it may be difficult to effectively distinguish subtle differences between DMUs with similar efficiencies (Wan et al., 2023) [4]. Moreover, the method does not fully account for the impact of random elements within the evaluation system, which could lead to deviations between the evaluation results and the actual situation in some cases. To address these issues, researchers have proposed to combine the bootstrap method with DEA to enhance the stability and reliability of the evaluation results, thereby providing a more accurate and reliable analytical tool for ecological efficiency research (Wan et al., 2023) [4]. The DEA model assumes precise input and output data, but in practice, data often contain errors and may be missing. To better evaluate industrial eco-efficiency and address the limitations of DEA in dealing with data uncertainty, many researchers have introduced the concept of fuzzy mathematics, with Zhou & Xu, (2020) [5] providing a comprehensive review.

The dynamic two-stage DDF model proposed by Zhuang et al. (2022) [6], when assessing the efficiency of urban water resource utilization and wastewater treatment in China, directly sets directional vectors without discussing how the choice of direction might affect the robustness of the results. Additionally, the model primarily focuses on technical efficiency, neglecting a comprehensive evaluation of economic and environmental efficiencies. Wan et al. (2023) [4] proposed a dynamic two-stage DEA model to assess the impact of industrial wastewater on public health in China. To address the assumption of unbiased data in DEA models and the potential measurement errors or biases in real-world corporate data, they employed the smoothed Bootstrap method. This technique uses resampling to identify and correct random errors and biases, thereby enhancing the accuracy of efficiency evaluations. These two articles did not adequately address the issues of data uncertainty and missing data in practical applications. Building upon them, this paper constructs a three-stage dynamic fuzzy DDF model, optimizes the smoothed bootstrap method, and introduces an endogenous directional vector into the DDF model. This model is not only applicable to water resource management but also extends its application to industrial production, environmental governance, and health protection within China's complex industrial ecosystem.

This paper optimizes the DEA model by integrating the proposed dynamic three-stage endogenous DDF model with the generalized smoothing bootstrap method and the LU partial order fuzzy comparison method, to create an integrated framework. This framework not only significantly enhances the precision of DEA evaluations but also broadens their application scope. The generalized smoothing bootstrap method offers confidence intervals for DEA efficiency scores, addressing several limitations of traditional smoothing bootstrap methods, such as their inability to handle non-normally distributed data. Furthermore, this method exhibits enhanced robustness in handling extreme values

and outliers, delivering more accurate estimation results. Its advantages are particularly pronounced when evaluating China's industrial ecosystem, where data distributions are complex and sample sizes are often limited. Introducing fuzzy comparison methods into the DEA model aims to address inaccuracies and missing data in China's industrial ecosystem. The LU partial order method represents uncertainty and fuzziness through the upper and lower bounds of fuzzy numbers, enabling an approximate representation of data ranges even when data is missing. By adopting this integrated framework, we address the limitations of traditional DEA in handling data uncertainty, correcting inherent biases, and enabling the model to adapt to the complexities of the China' ecological network system. The rest is structured as follows. Section 2 reviews the existing literature. Section 3 describes the methodology and data sources. Section 4 presents the results of the empirical study. Section 5 concludes the findings.

2. Literature Review

2.1. Bootstrap in DEA

As an effective multi-criteria efficiency evaluation method, DEA is capable of handling complex situations involving multiple inputs and outputs. It calculates the comparative efficiency of various DMUs, thereby providing decision-makers with an objective evaluation result. DEA is facing challenges related to of estimation bias. Moreover, the statistical properties of DEA estimators often prove intricate and difficult to directly address in practical applications. To confront these challenges, the bootstrapping method has emerged as an innovative means of inference. Not only does it effectively mitigate biases, but also it offers a more flexible solution. By augmenting inputs and outputs through the bootstrapping method, the constructed production frontier becomes more aligned with the complexities of the real world, significantly enhancing the accuracy of DEA models in assessing the efficiency of DMUs. The rigorous mathematical theories underlying the bootstrapping method have laid a solid foundation for its widespread application, ensuring stability and reliability in handling complex data (Simar & Wilson, 2015) [7].

Bootstrapping methods include smoothing bootstrap and subsampling bootstrap, both of which have been extensively applied in various fields. Smoothing bootstrap improves estimation by smoothing the edges of the data, while subsampling bootstrap performs inference by drawing smaller subsets from the original sample. Simar & Wilson (1998) [8] introduced a smoothing bootstrap method to provide confidence intervals and adjustments for inherent biases in the DEA process. Simar & Wilson (2000) [9] generalized the smoothing bootstrap to allow for heterogeneity in the efficiency structure, thereby enabling more flexible handling of potential differences among different DMUs. Kneip et al. (2008) [10] provided important theoretical foundations for DEA estimators. They explored the asymptotic distribution of DEA estimators and proved the consistency of both subsampling and smoothing bootstrap. Simar et al. (2012) [11] proposed a statistical inference method for DDFs, extending the asymptotic results of Kneip et al. (2008) [10] on radial DEA estimators. This extension enables the application of statistical inference to DEA estimators based on DDFs. Simar & Wilson (2011) [12] introduced a data-based algorithm to select an optimal subsample size, which identifies the most stable subsample size by calculating the variability under different subsample sizes. To tackle the uncertainties and randomness inherent in the evaluation process, Lin & Lu (2024) [13] compared the bootstrap method with chance-constrained DEA and stochastic nonparametric DEA, highlighting the significant advantages of the bootstrap approach.

When the data deviate from a normal distribution, the generalized smooth bootstrap method can provide more reasonable estimates. The flexibility of the generalized bootstrap method enables it to adapt to different data structures and model requirements. It can better

handle noise and outliers in the data, thereby improving the robustness of the estimates and helping calculate the confidence interval of the efficiency score, to make statistical inference more reliable.

2.2. Fuzzy Mathematics with DDF

The effectiveness of DEA largely hinges on the chosen models, and the incorporation of fuzzy mathematics concepts enables a more comprehensive capture of the critical factors influencing efficiency. Furthermore, DEA is heavily dependent on data, requiring a substantial amount of data to ensure result accuracy. Consequently, the quality and integrity of the data significantly impact the analysis outcomes. In contrast, the bootstrap method approximates the distribution of statistics through repeated resampling from the original dataset to generate pseudo-samples. This approach is particularly suitable for scenarios with small sample sizes or unknown data distributions. In the real world, data are often accompanied by uncertainties caused by measurement errors or data collection limitations. Fuzzy mathematics provides a framework to deal with such uncertainties. When confronted with missing or the incomplete data, fuzzy DEA can leverage available information for reasonable estimation, rather than simply ignoring incomplete data. In this way, fuzzy DEA demonstrates its unique advantages and value in handling complex and uncertain data environments.

Aparicio et al. (2023) [14] converted the fuzzy DEA model into a series of traditional DEA models using the α-level method proposed by Kao, (2000) [15], combining dynamic DDF and fuzzy logic to evaluate efficiency. Arana-Jiménez et al. (2022) [16] proposed a fuzzy DEA slacks-based approach, used to evaluate the relative efficiency of DMUs when input and output data are presented in the form of fuzzy sets. Mozaffari et al. (2021) [17] introduced a two-stage DEA network model embedded within a multi-objective linear programming framework to evaluate the efficiency of DMUs in a fully fuzzy environment, which incorporates fuzzy triangular numbers to simulate imprecise data, while also considering undesirable outputs. Zhang et al. (2017) [18] constructed a dynamic supply chain model based on Takagi-Sugeno fuzzy control theory, taking various factors into account, and validated the effectiveness of the model through simulation experiments. In complex decision-making environments, traditional decision-making methods often struggle to handle uncertainty and ambiguity. Huang et al. (2022) [19] proposed a veto multi-attribute grey target decision-making method, which provides effective solutions through a comprehensive effect measurement matrix. Li et al. (2024) [20] proposed the RFL-APIA framework, which enhances the performance of federated learning systems by introducing fuzzy rules, and the dynamic aggregation mechanism of this framework specifically takes into account the contribution of each client to the global model, thereby improving the robustness and efficiency of the overall model.

Different technologies or production processes may exhibit varying levels of efficiency and performance. Fuzzy DEA can address technological heterogeneity by providing a unified approach for comparison and evaluation. When assessing dynamic performance that evolves over time, fuzzy DEA can better capture uncertainties and variations in time-series data. The α-cut method and fuzzy ranking method are commonly used in handling fuzzy data comparisons (Hatami-Marbini et al., 2011) [21]. Using the α-cut method, fuzzy inputs and outputs can be converted into multiple α-level sets, and the DEA model is then applied independently to each α-level set. Consequently, evaluators can analyze the efficiency of DMUs at different confidence levels, yielding a series of efficiency scores instead of a single score. In contrast, the fuzzy ranking method does not directly convert fuzzy numbers into definite values or intervals; instead, it employs fuzzy relations or fuzzy logic to determine the relative efficiency among different DMUs, which does not result

in a precise numerical efficiency score. Zhou & Xu (2020) [5] provided a comprehensive overview of fuzzy DEA research from both theoretical and practical perspectives, detailing its successful applications in multiple fields.

This paper proposes a novel approach to handling fuzzy numbers based on Arana-Jiménez et al. (2022) [16], utilizing LU-fuzzy partial order (Wu, 2009; Stefanini, 2019) [22,23]. In fuzzy sets, the degree of membership of an element is no longer simply 0 or 1, but a continuous value between 0 and 1, indicating the extent to which the element belongs to a particular set. The LU-fuzzy partial order relation extends the traditional partial order relation by allowing for fuzziness when comparing elements. LU-fuzzy partial order quantifies the complex and interrelated factors in the IPEGHT industrial eco-system using fuzzy sets. It can address uncertainty through fuzzy logic to express ambiguous relationships between elements, thereby achieving comprehensive and accurate system evaluation.

2.3. Dynamic DEA Research on Eco-System

There exists a time lag in technological development and application, which traditional DEA models may not fully capture in terms of temporal dynamics. By unveiling the "black box" of traditional DEA models and considering the internal network structure of DMUs, this approach can not only evaluate the overall ecological efficiency of IPEGHT over the entire observation period, but also assess the dynamic changes in its periodic and stage efficiencies. This integrated approach, which combines vertical processing within the network structure with horizontal connections between successive periods, provides a more accurate reflection of the industrial ecological efficiency of each province (Tone & Tsutsui, 2010; 2014) [2,24].

Färe et al. (2013) [25] introduced the concept of "carry-over" assets to describe the transfer and utilization of resources across time periods in dynamic production. These assets represent intertemporal production factors, such as residual or reserved resources for future use, enabling the model to capture multi-period resource optimization and the utilization of surplus outputs, reflecting production decisions and efficiency. This not only increases the complexity of the model but also provides deeper insights into long-term strategic planning and resource management practices of enterprises. Fukuyama & Weber (2014) [26] proposed a dynamic two-stage network model, particularly focusing on the dynamic characteristics of non-performing loans and carryover assets. Fukuyama & Weber (2017) [27] further analyzed the impact of carry-over assets on the future production capacity of banks, highlighting the significance of carry-over assets in dynamic network DEA models and their influence on the production efficiency and performance of banks across different periods. Fukuyama et al. (2023) [28] explored the efficiency of China's banking sector through a dynamic network data envelopment analysis model, with a particular focus on carry-over assets—assets carried over between periods by banks—and their impact on the future production capacity of these banks. Considering dynamic DEA instead of static analysis allows for capturing dynamic changes, handling carry-over assets, reflecting long-term efficiency, accommodating complex production processes, and providing more comprehensive decision support, thereby significantly enhancing the accuracy and practicality of the model.

Industrial production is crucial to the economic prosperity of all nations, yet it simultaneously generates pollution that impacts human health and contributes to climate change (Sueyoshi et al., 2017) [29]. Therefore, it is imperative that we should strike a balance between pollution reduction and economic success in order to sustain global societal sustainability. Zhou et al. (2020) [30] conducted a spatial-temporal evolution analysis of ecological efficiency in 30 provinces of China from 2004 to 2016 using a three-stage DEA approach based on slack variables. By excluding the influences of external environments

and random errors, they found the ecological efficiency of most provinces is lower than the traditional estimated values. Zhang & Liu (2021) [31] considered the super-efficiency DEA and slack-based measurement approach for undesirable outputs, which not only enhances the accuracy of ecological efficiency measurement, but also aids enterprises in identifying and reducing resource consumption and pollutant emissions during the production process. Although Diao et al., (2021) [32] focused on performance evaluation through a decision tree model, their findings are closely related to the optimization of management efficiency, which aligns with the objectives of dynamic DEA. DEA, as a tool for assessing the relative efficiency of decision-making units, can be combined with decision tree models to provide a more comprehensive perspective on performance evaluation. Yang et al. (2024) [1] evaluated the dynamic relationship among agricultural and industrial economic growth, environmental pollution, and public health across 30 Chinese provinces from 2016 to 2020. By decomposing the overall efficiency, it clearly demonstrated the efficiency differences across different stages, providing a scientific basis for building a Healthy China and promoting high-quality development in both agricultural and industrial economies. Wang et al. (2021) [33] used a dynamic network DEA methodology to evaluate the two-stage ecological efficiency of China's industrial sector from 2010 to 2015, and analyzed the dynamic characteristics of capital investment across these two stages and their impact on ecological efficiency. Chen et al. (2024) [34] proposed a comprehensive artificial intelligence design approach that leverages Visual Question Answering technology to achieve deep integration of image and text information, aiming to fully understand and address environmental protection challenges, and provide innovative solutions. This method is highly consistent with the core objectives in the field of environmental management optimization in our current research.

Traditional DDF models often assume that input and output variables are exogenous, which means they are not controlled by the DMUs themselves. However, these variables may be influenced by the behaviors and decisions of the DMUs. Endogenous DDF models can provide a more accurate measurement of efficiency by considering these endogenous factors. Yang et al. (2018) [35] proposed the concept of "expected efficiency", and defined it as the average of a DMU's relative efficiency scores across all possible directions. This approach calculates the overall relative efficiency by considering all directions, overcoming subjectivity, and ensuring consistent and fair estimation for all DMUs. Song & Wang (2018) [3] proposed a slack-based endogenous DDF model, which selects and normalizes the directional vectors based on slack values, ensuring unit invariance and suitability of efficiency measurements for economic interpretation. Lozano & Soltani (2018) [36] used a nonlinear optimization model to determine a direction vector that calculates the minimum improvement required to reach the efficiency frontier. The fundamental idea behind this approach is to modify the maximum magnitude of improvement model or the unit-invariant version (Hampf & Krüger., 2015) [37], not by maximizing the improvement but by calculating the minimum improvement necessary to achieve efficiency.

Many researches have focused on the treatment technologies for the "three wastes" (waste water, waste gas, and solid waste) generated during industrial production processes(Shao et al., 2019) [38], but there is a lack of attention paid to their potential impacts on the environment and human health. Although preliminary investigations have been conducted (Zhang et al., 2021; Wan et al., 2023) [4,31], this issue still needs to be in-depth analyzed and researched. Future efforts should strengthen the assessment of the impacts of industrial pollution on public health to promote sustainable industrial development and environmental protection.

2.4. Summary of Literature

Table 1 shows the methodologies employed in this study, while drawing insights and making comparisons with the existing literature. Hampf & Krüger (2015) [36] considered the tradeoff between increased output and reduced emissions, and applied optimization techniques to determine the optimal direction for DMUs in environmental efficiency analysis. Arana-Jiménez et al. (2022) [16] extended the traditional DEA slacks method by introducing fuzzy mathematical theory to handle the uncertainty in data. Shabani et al. (2023) [39] applied the methodology of Arana-Jiménez et al. (2022) [16] to dynamic network systems. Zhuang et al. (2022) [6] utilized the dynamic network DDF model for evaluating wastewater treatment. Wan et al. (2023) [4] employed a refined smoothing bootstrap approach within the context of assessing the potential health risks associated with industrial wastewater.

Table 1. Literature review.

Method Paper	DEA Model	DDF Model	Bootstrap In DEA	Endogenous Vector	Network Structure	Dynamic Assessment	Fuzzy DEA
Hampf & Krüger (2015) [36]	√	√		√			
Zhuang et al. (2022) [6]	√	√			√	√	
Arana-Jiménez et al. (2022) [16]	√				√		√
Shabani et al. (2023) [39]	√	√			√	√	√
Wan et al. (2023) [4]	√		√		√	√	
This paper	√	√	√	√	√	√	√

Due to inherent biases in traditional DEA and its inability to handle uncertainty, this paper introduces the generalized smooth bootstrap method and fuzzy mathematics. Based on the static network DEA method, this paper considers the dynamic characteristics of capital investment across different time periods, considering not only the impact of current period's capital investment on eco-efficiency but also how these investments continue to play a role in subsequent periods. To achieve this goal, the DDF model is adopted, allowing for the definition of different measurement directions for each input or output. Some researchers directly used the observed values of DMUs as direction vectors, which may lead to an overestimation of efficiency values. Furthermore, previous methods for determining endogenous direction vectors were often intricate and laden with numerous constraints. Consequently, this study initially adopts a simpler approach by using observed values as the direction vectors, while implementing corrective measures through the application of weight coefficients, α and β, to refine the analysis.

In multi-stage systems, the contributions of different stages may vary significantly, and traditional fixed-weight allocation methods often fail to accurately reflect the actual contributions of each stage. To address this, the SoftMax function can be employed to dynamically calculate weights based on the inputs of each stage, thereby achieving more precise weight allocation. The output of the SoftMax function is a normalized probability distribution, ensuring that the sum of all weights equals 1. This characteristic not only guarantees reasonable weight allocation but also effectively avoids issues of uneven weight distribution, significantly enhancing the accuracy and rationality of the evaluation results. Wu et al. (2024) [40] proposed an improved softmax method that is superior to the traditional SoftMax.

The pollution issues caused by industrial production remain a significant problem. Implementing stringent standards for the treatment of "three wastes" is not only a necessary measure to protect the ecological environment, but also a key to achieving harmonious coexistence between economic sustainability and environmental protection. The quality of the environment directly impacts human health, and the government should significantly increase resource allocation in the health sector. This aims to reduce pollutant emissions and promote the transition to green production methods through multi-dimensional approaches

including technological innovation, policy guidance, and public participation, thereby expediting the realization of SDGs.

This paper employs an integrated approach aimed at enhancing the accuracy and broad applicability of model evaluation. Initially, the proposed endogenous DDF model is used to obtain the initial efficiency. In the second step, the generalized smoothed bootstrap method is applied to correct the initial efficiency; however, the results of the bootstrap method are contingent upon the representativeness of the original sample. If the sample is biased or not representative, the bootstrap method may fail to provide accurate inferences. Consequently, in the third step, fuzzy mathematics is introduced, employing the LU-fuzzy partial order to compare fuzzy numbers and derive the results. A simplified approach is proposed to select endogenous directional vectors, aiming to streamline the process while maintaining the objectivity and reliability of the evaluation results. This method comprehensively considers various potential factors, minimizes the constraints on the model, and maintains the simplicity of the calculations, enabling this integrated scheme to be flexibly applied across multiple domains and complex scenarios. The main contributions of this paper are as follows: (1) A dynamic three-stage DDF model, integrating the generalized smooth bootstrap method and fuzzy mathematics, is proposed. (2) A straightforward method for selecting endogeneity direction vectors is proposed, which guides efficiency improvements for each province. (3) This paper applies the proposed model and methods to the IPEGHT industrial eco-system, which operates in a dynamic three-stage environment.

3. Methodology and Data

3.1. Dynamic Endogenous DDF Model

Figure 1 illustrates the integrated framework proposed in this paper for calculating IPEGHT efficiency. This framework is a dynamic three-stage bootstrap model that incorporates an endogenous fuzzy directional distance function (DDF). The specific process is as follows: first, fuzzifying the data to determine the upper and lower bounds of the fuzzy numbers; then, using Equation (5) to calculate the precise DDF model; next, introducing the LU partial order fuzzy comparison method to derive Equation (10), which is used to calculate the fuzzy numbers; finally, combining the generalized smooth bootstrap method to perform 2000 resamples, generating 2000 samples. Each sample is iteratively input into Equation (10) to calculate the efficiency values, compute the mean, and correct the original efficiency, resulting in the final outcome.

In response to the SDGs, China, as a populous and resource-intensive country, bears significant responsibility and obligation. As depicted in Figure 1, the three critical stages of IP, EG, and HT together form a comprehensive and intricate eco-system. While the IP stage has propelled economic development, it has also given rise to issues of "three wastes". The EG stage is dedicated to addressing these environmental pollutants. However, the HT stage reveals that the residual "three wastes" emitted during the EG stage have caused environmental contamination, which in turn has impacted human health. These three stages have a progressive relationship, especially in China, where environmental protection still faces many challenges and requires urgent improvement.

In period t, $x_{ijt}^{IP}(t=1,\cdots,T; i=1,2; j=1,\cdots,n)$ and $y_{rjt}^{IP} \in R^{+}$ $(t=1,\cdots,T; r=1; j=1,\cdots,n)$ are defined as the input i and output r of DMU_j in the IP stage, respectively; similarly, x_{ijt}^{EG} and y_{rjt}^{EG} are defined in the EG stage; similarly, x_{ijt}^{HT} and are defined in the HT stage. $z_{jt}^{L1} \in R^{+}(t=1,\cdots,T; j=1,\cdots n)$ is defined as a connection for IP stage and EG stage, $z_{jt}^{L2} \in R^{+}(t=1,\cdots,T; j=1,\cdots n)$ is defined as a connection for EG stage and HT stage, $z_{j(t,t+1)}^{C} \in R^{+}(t=1,\cdots,T-1; j=1,2,\cdots,n)$ is characterized

as a good carry-over, transferring from the IP stage to the corresponding stage of DMU_j between consecutive periods t and $t + 1$.

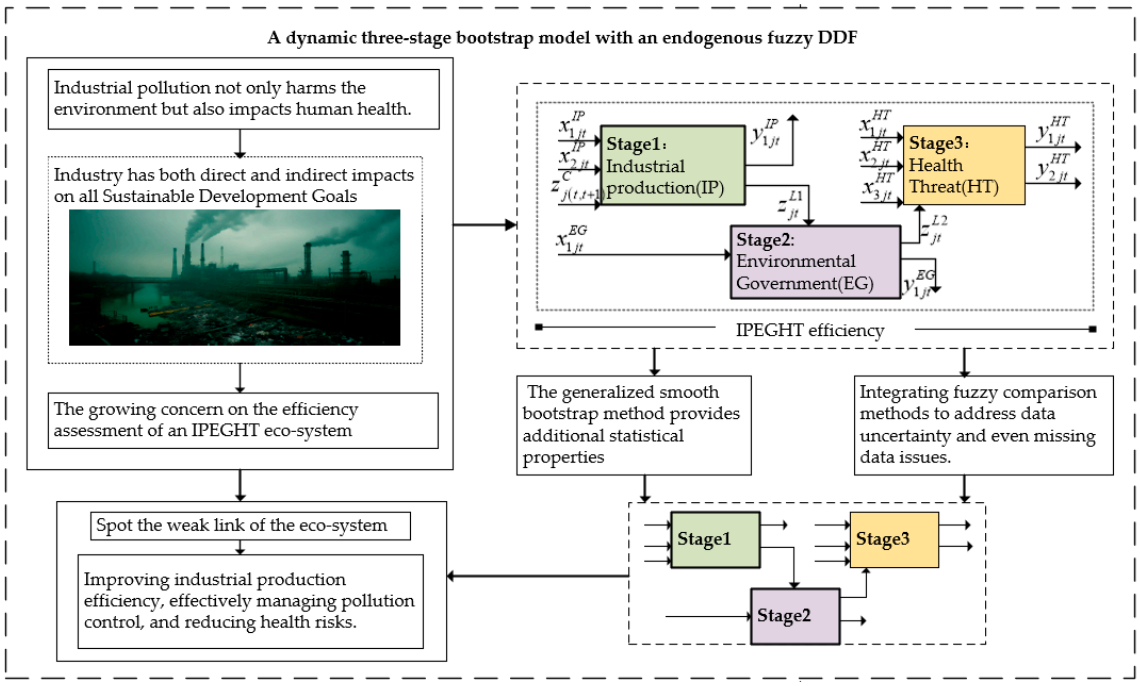


Figure 1. A dynamic three-stage model with endogenous DDF for IPEGHT system.

Table 2 shows the indicator variables in the IPEGHT eco-system. The "three wastes" production index is derived by summing the standardized values of chemical oxygen demand from wastewater production, sulfur dioxide generation, and general industrial solid waste production. The "three wastes" emission index is calculated by summing the standardized values of chemical oxygen demand from wastewater discharge, sulfur dioxide emissions, and general industrial solid waste discharge.

Table 2. Variables in the IPEGHT system.

Stage		Variables	Units
Stage 1: IP	input	Energy consumption	Ten thousand tons
		Industrial employees	million persons
	output	Industrial added assets	billion CNY
stage link 1		Three wastes production index	-
stage 2: EG	input	Completed investment in industrial pollution control	Ten thousand CNY
	output	Comprehensive value of three wastes reuse	Ten thousand CNY
stage link 2		Three wastes emission index	-
stage 3: HT	input	number of health technicians	per 1000 population
		number of healthcare beds	per 1000 individuals
		Per capita health expenditure	CNY
	output	Average number of medical visits by residents	-
		Health index	-
period link	carry-over	industrial fixed assets	hundred million CNY

3.2. The Generalized Smooth Bootstrap Method

Based on the work of Simar & Wilson (2000) [9], the algorithm for the generalized smooth bootstrap method is summarized as follows:

Step 1. The dataset D is used to solve $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n$ with Equation (5).

Step 2. Transform coordinate, $(x_i, y_i) \rightarrow (y_i, \eta_i, \hat{\theta}_i) \forall i = 1, 2, \cdots, n$.

Step 3. $Z = [y_i, \eta_i, \hat{\theta}_i]$ is the defined matrix of $n \times (p+q)$ and $Z_R = [y_i, \eta_i, 2 - \hat{\theta}_i]$ is reflected about boundary $\hat{\theta}_i = 1$. $\hat{Z} = \begin{pmatrix} Z \\ Z_R \end{pmatrix}$ is the augmented matrix. Calculate $Z, Z_R, \hat{Z}$. $\hat{\Sigma}_1, \hat{\Sigma}_2$ are covariance estimates of Z, Z_R, respectively. Using Cholesky decomposition $\hat{\Sigma}_1 = L_1 L_1^T, \hat{\Sigma}_2 = L_2 L_2^T$.

Step 4. n row vectors are randomly selected from the $\hat{Z}$ to form $n \times (p+q)$ matrix $\hat{Z}^*$. Calculate the average of each column of $\hat{Z}^*$ to obtain the vector $\overline{z}^*$ of $1 \times (p+q)$.

Step 5. Choose an appropriate bandwidth h.

Step 6. Randomly generate a *i.i.d.* matrix ε of $n \times (p+q)$, which is standard normal pseudo-random variates. ε_i is ith row of ε. Compute $n \times (p+q)$ matrix ε^* with the ith row $\varepsilon_{i\cdot}^*$ given by $\varepsilon_{i\cdot}^* = \varepsilon_{i\cdot} L_l^T, l = 1, 2$.

Step 7. Calculate matrix $\Gamma = \frac{M\hat{Z}^* + h\varepsilon^*}{\sqrt{1+h^2}} + i_n \otimes \overline{z}^*$ of $n \times (p+q)$ where $M = I_n - \frac{i_n i_n^T}{n}$ with I_n is the identity matrix of order n and i_n is an $n \times 1$ vector of ones, and $\otimes$ is the Kronecker product.

Step 8. $\Gamma = [\gamma_{i1}, \gamma_{i2}, \gamma_{i3}] . \gamma_{i1} \in \mathbb{R}_+^q, \gamma_{i2} \in [0, \pi/2]^{p-1}$ and $\gamma_{i3} \in \mathbb{R}$ $\forall i = 1, 2, \cdots, n$. Denote the $n \times (p+q)$ matrix of bootstrap pseudo-sample Z^* such that the ith row z_i^* of Z^* is given by $z_i^* = \begin{cases} (\gamma_{i1}, \gamma_{i2}, \gamma_{i3}) if \gamma_{i3} \geq 1 \\ (\gamma_{i1}, \gamma_{i2}, 2 - \gamma_{i3}) if \gamma_{i3} < 1 \end{cases}$.

Step 9. Transform polar coordinates in Z^* to Cartesian coordinates using the formula $x_i^* = \hat{x}\hat{\theta}_i^* / \hat{\theta}$; calculate a pseudo-sample D^*, where $x_{bijt}^{*IWU} = \left(\hat{\theta}_{jb}^* / \hat{\theta}_j\right) x_{ijt}^{IWU}$, and $x_{bijt}^{*HR} = \left(\hat{\theta}_{jb}^* / \hat{\theta}_j\right) x_{ijt}^{HR}, j = 1, 2, \cdots, n$.

Step 10. Calculate $\hat{\theta}_{jb}^*$ of $\hat{\theta}_j$ for $j = 1, \cdots, n$ by Equation (5) with D^*.

Step 11. Repeat step 5 to 10 for B = 2000 iterations to generate a collection of estimations $\left\{\hat{\theta}_{jb}^*, b = 1, \cdots, B; j = 1, \cdots, n\right\}$ that ensure sufficient coverage of the 95% confidence intervals. Calculate α confidence interval $\theta_j \in [\hat{\theta}_j + \hat{a}_\alpha, \hat{\theta}_j + \hat{b}_\alpha]$ by $P\left(-\hat{b}_\alpha \leq \hat{\theta}_{jb}^* - \hat{\theta}_j \leq -\hat{a}_\alpha\right) = 1 - \alpha$, $P\left(-\hat{b}_\alpha \leq \hat{\theta}_j - \theta_j \leq -\hat{a}_\alpha\right) = 1 - \alpha$.

3.3. The Fuzzy Method

This study applies fuzzification to the data sourced from statistical yearbooks. The trapezoidal fuzzy number is employed as fuzzification tool, with the aim of more precisely capturing and articulating the intricate relationships within the data. The application of trapezoidal fuzzy numbers enhances the robustness of the model, enabling it to effectively handle the ambiguity and uncertainty inherent in real-world problems.

Definition 1. *A trapezoidal fuzzy number is typically represented as a quadruple* (a_1, a_2, a_3, a_4)*, with its membership function defined as Equation (1) (Nadarajan, Ahmed et al., 2023) [41].*

$$\mu(x) = \begin{cases} \frac{x - a_1}{a_2 - a_1}, a_1 \leq x < a_2 \\ 1, a_2 \leq x \leq a_3 \\ \frac{a_4 - x}{a_4 - a_3}, a_3 < x \leq a_4 \\ 0, otherwise \end{cases} \quad (1)$$

The membership function indicates that values within the interval $[a_2, a_3]$ possess full membership, with a degree of 1. Meanwhile, within the intervals $[a_1, a_2]$ and $[a_3, a_4]$, the degree of membership gradually increases or decreases. When a_2 equals a_3, the trapezoidal

fuzzy number transforms into a triangular fuzzy number. It is important to note that the data in this study are all non-negative.

Definition 2. *$\widetilde{A}$ is the trapezoidal fuzzy number, whose membership function exhibits the following characteristics, as referenced from Gerami et al. (2023) [42]:*

1. *$\widetilde{A}$ is convex, i.e., $\mu(x_1) \geq \mu(x_2), \forall x \in (x_1, x_2), \mu(x) \geq min(\mu(x_1), \mu(x_2))$*
2. *$\widetilde{A}$ is normal, i.e., $\exists x \in R, \mu_{\widetilde{A}}(x) = 1$*
3. *$\mu_{\widetilde{A}}(x)$ is piecewise continuous.*
4. *$core(\widetilde{A}) = \{x \in R | \mu_{\widetilde{A}}(x) = 1\}$*

The $core(\widetilde{A})$ encompasses all values with the highest degree of membership, commonly regarded as the most probable range of values.

Definition 3. *The α-cut is the subset of elements in the fuzzy set where the membership degree is greater than or equal to a certain threshold α. For a fuzzy number $\widetilde{A}$ and a given α value, the α-cut can be expressed as: $\widetilde{A}_\alpha = \{x \in X | \mu_{\widetilde{A}}(x) \geq \alpha\}$. Its corresponding α-levels are determined by $\widetilde{A}_\alpha = [a_1^\alpha, a_2^\alpha]$, $a_2^\alpha = a_4 + \alpha(a_4 - a_3)$, $a_1^\alpha = a_1 + \alpha(a_2 - a_1)$ (Arana-Jiménez et al., 2022) [16].*

Definition 4. *LU-fuzzy partial orders are a type of partial order relation defined within the framework of fuzzy set theory, which combines fuzzy logic and partial order theory. They establish an ordering relationship by comparing the α-cut of numbers. P and q are two non-negative numbers, $p \leq q$ if and only if $\underline{p}_\alpha \leq \underline{q}_\alpha$ and $\overline{p}_\alpha \leq \overline{q}_\alpha$, for all $a \in [0, 1]$. For fuzzy numbers $\widetilde{b}$ and $\widetilde{c}$, $\widetilde{b} \leq \widetilde{c}$ if and only if $\widetilde{b}_\alpha \leq \widetilde{c}_\alpha$ for all $\alpha \in [0, 1]$ (Arana-Jiménez et al., 2022) [16].*

Definition 5. *$\widetilde{b} = (b_1, b_2, b_3, b_4)$ and $\widetilde{c} = (c_1, c_2, c_3, c_4)$ are two non-negative trapezoidal fuzzy numbers and $\lambda \in R^+$, as referenced from Gerami et al. (2023) [42], then*

$$\widetilde{b} \pm \widetilde{c} = (b_1 \pm c_1, b_2 \pm c_2, b_3 \pm c_3, b_4 \pm c_4) \tag{2}$$

$$\lambda\widetilde{b} = (\lambda b_1, \lambda b_2, \lambda b_3, \lambda b_4) \tag{3}$$

$$\widetilde{b} \otimes \widetilde{c} = (b_1 \times c_1, b_2 \times c_2, b_3 \times c_3, b_4 \times c_4) \tag{4}$$

Based on the definition, the LU-fuzzy partial orders used in this study can be extended to trapezoidal fuzzy numbers by leveraging the characterization result given by triangular fuzzy numbers (Arana-Jiménez et al., 2022) [16]. $\widetilde{p} = (p_1, p_2, p_3, p_4)$ and $\widetilde{q} = (q_1, q_2, q_3, q_4)$ are two non-negative trapezoidal fuzzy numbers, $\widetilde{p} \leq \widetilde{q}$ if and only if $p^i \leq q^i$, for all $i = 1, 2, 3, 4$.

This study employs an endogenous directional vector, adjusting based on an exogenous directional vector. Specifically, an initial directional vector is assigned to DMUs, and subsequently, uncertain parameters are introduced to traverse and identify the direction that maximizes the sum of all potential output improvements. The initialization directional vector is selected from observed values, with undesirable outputs taken into consideration.

By obtaining the optimal inputs, it can determine the potential adjustment space for both desirable and undesirable outputs. In this three-stage intricate network structure, the SoftMax function $w_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_i}}$ is employed to dynamically adjust stage weights endogenously, thereby addressing the issue of weight allocation and ensuring a rational evaluation of eco-system efficiency scores (Nitta, 2003) [43]. Wu et al. (2024) [40] proposed an improved softmax method that is superior to the traditional SoftMax.

Based on the discussion above, the model proposed in this paper is a three-stage dynamic DDF model that integrates fuzzy logic and bootstrapping methods. The objective function is as Equation (5).

$$\max \sum_{t=1}^{T} \gamma_t (w^{IP,t}\widetilde{\theta}_s^{IP,t} + w^{EG,t}\widetilde{\theta}_s^{EG,t} + w^{HT,t}\widetilde{\theta}_s^{HT,t}) \tag{5}$$

The constraints are as Equations (6)–(9). The first three represent the individual constraints for the IP, EG, and HT stage, respectively. Equation (9) serves as a crucial constraint that governs the inter-stage connections and carry-over effects.

$$\begin{cases} \sum\limits_{j=1}^{n} \widetilde{x}_{ji}^{IP,t}\lambda_j^{IP,t} \le (1-\widetilde{\theta}_s^{IP,t})\widetilde{x}_{j'i}^{IP,t} \; s=1,2,3,4; i=1,\cdots I \\ \sum\limits_{j=1}^{n} \widetilde{y}_{jk}^{IP,t}\lambda_j^{IP,t} \ge (1+\alpha_{j'k}^{IP,t}\widetilde{\theta}_s^{IP,t})\widetilde{y}_{j'k}^{IP,t} \; s=1,2,3,4; k=1,\cdots,K \\ \sum\limits_{j=1}^{n} \widetilde{b}_{jm}^{IP,t}\lambda_j^{IP,t} \le (1-\widetilde{\beta}_{j'm}^{IP,t}\widetilde{\theta}_s^{IP,t})\widetilde{b}_{j'm}^{IP,t} \; s=1,2,3,4; m=1,\cdots,M \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{IP} = 1 \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{IP,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{IP,t} = 1 \\ \alpha_{j'k}^{IP,t}, \beta_{j'm}^{IP,t} \ge 0 \quad k=1,\cdots,K; \; m=1,\cdots,M; t=1,\cdots,T \end{cases} \tag{6}$$

$$\begin{cases} \sum\limits_{j=1}^{n} \widetilde{x}_{ji}^{EG,t}\lambda_j^{EG,t} \le (1-\widetilde{\theta}_s^{EG,t})\widetilde{x}_{j'i}^{EG} \; s=1,2,3,4; i=1,\cdots I \\ \sum\limits_{j=1}^{n} \widetilde{y}_{jk}^{EG,t}\lambda_{jt}^{EG,t} \ge (1+\alpha_{j'k}^{EG,t}\widetilde{\theta}_s^{EG,t})\widetilde{y}_{j'k}^{EG,t} \; s=1,2,3,4; k=1,\cdots,K \\ \sum\limits_{j=1}^{n} \widetilde{b}_{jm}^{EG,t}\lambda_j^{EG,t} \le (1-\beta_{j'm}^{EG,t}\widetilde{\theta}_s^{EG,t})\widetilde{b}_{j'm}^{EG,t} \; s=1,2,3,4; m=1,\cdots,M \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{EG} = 1 \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{EG,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{EG,t} = 1 \\ \alpha_{j'k}^{EG}, \beta_{j'm}^{EG} \ge 0 \quad k=1,\cdots,K; \; m=1,\cdots,M; t=1,\cdots,T \end{cases} \tag{7}$$

$$\begin{cases} \sum\limits_{j=1}^{n} \widetilde{x}_{ji}^{HT,t}\lambda_j^{HT,t} \le (1-\widetilde{\theta}_s^{HT,t})\widetilde{x}_{j'i}^{HT} \; s=1,2,3,4; i=1,\cdots I \\ \sum\limits_{j=1}^{n} \widetilde{y}_{jk}^{HT,t}\lambda_{jt}^{HT,t} \ge (1+\alpha_{j'k}^{HT,t}\widetilde{\theta}_s^{HT,t})\widetilde{y}_{j'k}^{HT,t} \; s=1,2,3,4; k=1,\cdots,K \\ \sum\limits_{j=1}^{n} \widetilde{b}_{jm}^{HT,t}\lambda_j^{HT,t} \le (1-\beta_{j'm}^{HT,t}\widetilde{\theta}_s^{HT,t})\widetilde{b}_{j'm}^{HT,t} \; s=1,2,3,4; m=1,\cdots,M \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{HR} = 1 \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{HT,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{HT,t} = 1 \\ \alpha_{j'k}^{HT,t}, \beta_{j'm}^{HT,t} \ge 0 \quad k=1,\cdots,K; \; m=1,\cdots,M; t=1,\cdots,T \end{cases} \tag{8}$$

$$\begin{cases} \left.\sum\limits_{j=1}^{n} \widetilde{Z}_{jt}^{L_1}\lambda_{jt}^{IP} = \sum\limits_{j=1}^{n} \widetilde{Z}_{jt}^{L_1}\lambda_{jt}^{EG}, \right\} stage\ link1 \\ \left.\sum\limits_{j=1}^{n} \widetilde{Z}_{jt}^{L_2}\lambda_{jt}^{EG} = \sum\limits_{j=1}^{n} \widetilde{Z}_{jt}^{L_2}\lambda_{jt}^{HR}, \right\} stage\ link2 \\ \left.\sum\limits_{j=1}^{n} \widetilde{Z}_{j(t,t+1)}^{c}\lambda_{jt}^{IP} = \sum\limits_{j=1}^{n} \widetilde{Z}_{j(t,t+1)}^{c}\lambda_{j,t+1}^{IP}, \right\} Period\ link \end{cases} \tag{9}$$

Assuming all inputs and outputs are trapezoidal fuzzy numbers, the objective function likewise takes the form of a trapezoidal fuzzy number. The intensity variables $\lambda_{jt}^{v}, v = IP, EG, HT$ are non-negative real numbers. Fuzzy Pareto solutions are a method used in multi-objective optimization problems to find the optimal solutions under such fuzzy constraints. Based on the definitions provided above regarding fuzzy numbers, the following equivalent formula is derived for ease of calculation:

$$\max \sum_{t=1}^{T} \gamma_t \sum_{s=1}^{4} \frac{1}{4} (w^{IP,t} \theta_s^{IP,t} + w^{EG,t} \theta_s^{EG,t} + w^{HT,t} \theta_s^{HT,t}) \tag{10}$$

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ji}^{IP,t} \lambda_j^{IP,t} \leq (1 - \theta_s^{IP,t}) x_{j'i}^{IP,t} \; s = 1,2,3,4; i = 1, \cdots I \\ \sum\limits_{j=1}^{n} y_{jk}^{IP,t} \lambda_j^{IP,t} \geq (1 + \alpha_{j'k}^{IP,t} \theta_s^{IP,t}) y_{j'k}^{IP,t} \; s = 1,2,3,4; k = 1, \cdots, K \\ \sum\limits_{j=1}^{n} b_{jm}^{IP,t} \lambda_j^{IP,t} \leq (1 - \beta_{j'm}^{IP,t} \theta_s^{IP,t}) b_{j'm}^{IP,t} \; s = 1,2,3,4; m = 1, \cdots, M \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{IP,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{IP,t} = 1 \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{IP} = 1 \\ \alpha_{j'k}^{IP,t}, \beta_{j'm}^{IP,t} \geq 0 \quad k = 1, \cdots, K; \; m = 1, \cdots, M; t = 1, \cdots, T \\ \theta_t^s \leq \theta_t^{s+1} \; s = 1,2,3 \end{cases} \tag{11}$$

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ji}^{EG,t} \lambda_j^{EG,t} \leq (1 - \theta_s^{EG,t}) x_{j'i}^{EG} \; s = 1,2,3,4; i = 1, \cdots I \\ \sum\limits_{j=1}^{n} y_{jk}^{EG,t} \lambda_{jt}^{EG,t} \geq (1 + \alpha_{j'k}^{EG,t} \theta_s^{EG,t}) y_{j'k}^{EG,t} \; s = 1,2,3,4; k = 1, \cdots, K \\ \sum\limits_{j=1}^{n} b_{jm}^{EG,t} \lambda_j^{EG,t} \leq (1 - \beta_{j'm}^{EG,t} \theta_s^{EG,t}) b_{j'm}^{EG,t} \; s = 1,2,3,4; m = 1, \cdots, M \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{EG} = 1 \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{EG,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{EG,t} = 1 \\ \alpha_{j'k}^{EG}, \beta_{j'm}^{EG} \geq 0 \; k = 1, \cdots, K; \; m = 1, \cdots, M; t = 1, \cdots, T \\ \theta_t^s \leq \theta_t^{s+1} \; s = 1,2,3 \end{cases} \tag{12}$$

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ji}^{HT,t} \lambda_j^{HT,t} \leq (1 - \theta_s^{HT,t}) x_{j'i}^{HT} \; s = 1,2,3,4; i = 1, \cdots I \\ \sum\limits_{j=1}^{n} y_{jk}^{HT,t} \lambda_{jt}^{HT,t} \geq (1 + \alpha_{j'k}^{HT,t} \theta_s^{HT,t}) y_{j'k}^{HT,t} \; s = 1,2,3,4; k = 1, \cdots, K \\ \sum\limits_{j=1}^{n} b_{jm}^{HT,t} \lambda_j^{HT,t} \leq (1 - \beta_{j'm}^{HT,t} \theta_s^{HT,t}) b_{j'm}^{HT,t} \; s = 1,2,3,4; m = 1, \cdots, M \\ \sum\limits_{j=1}^{n} \lambda_{jt}^{HR} = 1 \\ \sum\limits_{k=1}^{K} \alpha_{j'k}^{HT,t} + \sum\limits_{m=1}^{M} \beta_{j'm}^{HT,t} = 1 \\ \alpha_{j'k}^{HT,t}, \beta_{j'm}^{HT,t} \geq 0 \; k = 1, \cdots, K; \; m = 1, \cdots, M; t = 1, \cdots, T \\ \theta_t^s \leq \theta_t^{s+1} \; s = 1,2,3 \end{cases} \tag{13}$$

$$\begin{cases} \left.\sum_{j=1}^{n} Z_{jt}^{L_1} \lambda_{jt}^{IP} = \sum_{j=1}^{n} Z_{jt}^{L_1} \lambda_{jt}^{EG},\right\} stage\ link1 \\ \left.\sum_{j=1}^{n} Z_{jt}^{L_2} \lambda_{jt}^{EG} = \sum_{j=1}^{n} Z_{jt}^{L_2} \lambda_{jt}^{HR},\right\} stage\ link2 \\ \left.\sum_{j=1}^{n} Z_{j(t,t+1)}^{c} \lambda_{jt}^{IP} = \sum_{j=1}^{n} Z_{j(t,t+1)}^{c} \lambda_{j,t+1}^{IP},\right\} Period\ link \end{cases} \tag{14}$$

The stage efficiency score is calculated using Equation (15).

$$\rho = max\left(1 - \sum_{t=1}^{T} \gamma_t \frac{1}{4} \sum_{s=1}^{4} \theta_s^t\right) \tag{15}$$

The efficiency score for period t is calculated using Equation (16).

$$\rho_t = \frac{1}{4} max\left(w^{IP,t} \times \sum_{s=1}^{4} \theta_s^{IP,t} + w^{EG,t} \times \sum_{s=1}^{4} \theta_s^{EG,t} + w^{HT,t} \times \sum_{s=1}^{4} \theta_s^{HT,t}\right) \tag{16}$$

The efficiency score of the system can also be obtained through the weighted sum of the efficiency scores for each period t, as Equation (17).

$$\rho = \sum_{t=1}^{T} \gamma_t \rho_t \tag{17}$$

This paper employs the SoftMax function to dynamically adjust the stage weights endogenously. By avoiding the use of weights, subjectivity is minimized, allowing for a more objective and genuine analysis of the dynamic evolution over the 11 years. Additionally, in multi-stage DEA, weights and efficiency scores are endogenous, and the assumptions about decomposition weights in the model can influence the overall efficiency assessment.

3.4. Data

The data of 30 provinces in Chinese mainland (excluding Tibet) are analyzed from 2011 to 2021, focusing on the dynamic changes in the efficiency of the IPEGHT eco-system. The dataset is primarily sourced from China Population and Employment Statistics Yearbook (NBSPES, 2022) [44], China Health Statistical Yearbook (NHC, 2022) [45], China Statistical Yearbook (NBSC, 2022) [46], China Environmental Statistics Yearbook (NBSCES, 2022) [47] as well as the respective provincial statistical yearbooks. To account for price fluctuations and ensure temporal comparability of industrial data, this study adopts a 2011-based price deflation method for value-added and fixed asset investment.

The generalized smoothing bootstrap algorithm does not rely on data distribution assumptions and can handle non-normal distribution data, such as skewed distributions or outliers. By introducing smoothing noise to simulate data fluctuations, it enhances the robustness of estimates. However, although it can handle different scales through data standardization, if the scale differences are too large, the algorithm may still struggle to fully eliminate their impact.

This paper innovatively integrates fuzzy comparison methods into the DEA model, effectively addressing challenges related to data uncertainty and even missing data. Although the data obtained in this study are presented in the form of precise values, considering the potential inaccuracies that may be inherent in these data, we have adopted a strategy that involves converting data x into trapezoidal fuzzy numbers $[0.9x, 0.95x, 1.05x, 1.1x]$. This transformation process not only enables a more accurate capture of the uncertainty characteristics of the data, but also upgrades the traditional DEA model to a fuzzy DEA model. When dealing with the indicator "Comprehensive Value of Three Wastes Reuse",

data for some provinces were missing, and linear interpolation was used for filling. This paper introduces the LU partial order fuzzy comparison method into the DEA model, which can address data missing issues to some extent but requires domain knowledge and experience to determine the upper and lower bounds of fuzzy numbers. Precise values of 0.9 and 1.1 were selected as the bounds, unavoidably introducing some subjectivity due to the reliance on predefined thresholds. However, using the LU partial order method to process large datasets may significantly reduce computational efficiency. Moreover, this method may not be suitable if there is a significant amount of missing data.

4. Empirical Analysis

4.1. Comparison of Endogenous Methods

Figure 2 illustrates a comparison of the IPEGHT efficiency obtained by using observed values as exogenous direction vectors versus the endogenously derived direction vectors proposed in this paper. "Observed values" represents the direct use of observed values of DMUs as directional vectors, with the actual performance of these provinces directly serving as a reference for efficiency evaluation. Most DDF models directly employ this method, such as Zhuang et al. (2022) [6].However, this approach is simplistic and may lead to an overestimation of efficiency values, as it fails to account for potential areas of improvement. In contrast, "endogenous direction vectors" represents the endogenous directional vector proposed in this paper, which implies that in evaluating efficiency, the comprehensive performance of all DMUs is taken into consideration, rather than the performance of each unit individually. Specifically, the endogenous direction vector is derived by comprehensively analyzing the input and output data of all DMUs, resulting in a direction vector that reflects the overall system characteristics. This method avoids the bias of exogenous direction vectors and reveals the interrelationships between units, providing a scientific basis for optimization decisions. This method more accurately reflects the efficiency of each province relative to the overall average level, as it considers the optimal performance of all provinces rather than relying solely on data from individual provinces. In Figure 2, it is evident that the use of observed values as direction vectors significantly outperforms the endogenous direction vectors. This finding further substantiates the notion that employing exogenous direction vectors may overestimate the actual efficiency of DMUs.

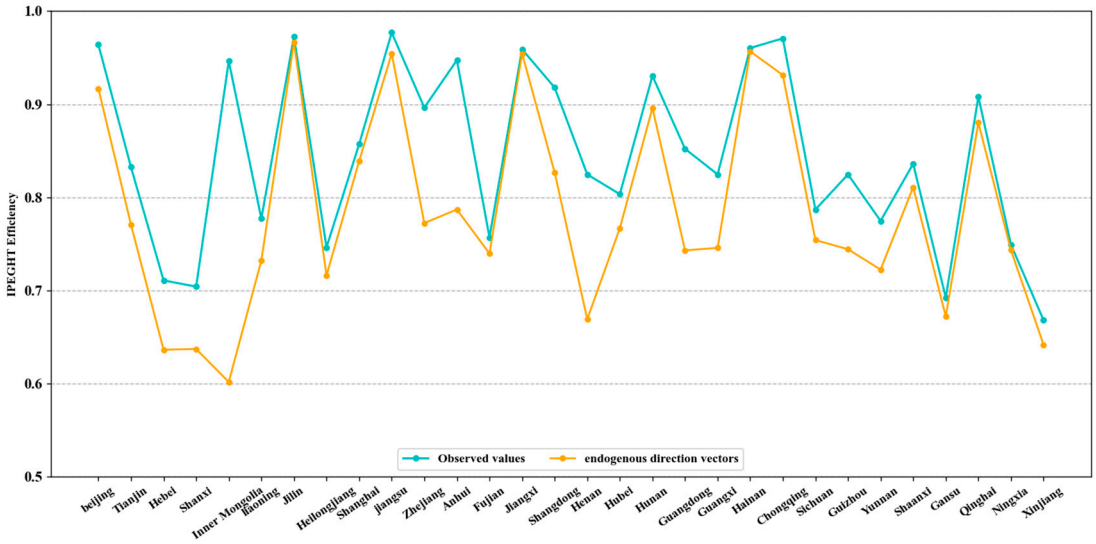


Figure 2. Observed values vs. endogenous direction vectors.

The endogenous direction vectors proposed in this paper can provide guidance for the future improvement of industrial ecological development in each province. Table 3 presents the values of the expected output weight α for the industrial eco-system during the IP stage. Given that the IP stage involves only one desirable output and one undesirable output, both α and β have a single value. It is important to note that the number of expected weights corresponds one-to-one with the number of expected outputs, and the constraint $\alpha_{j'k}^{IP,t} + \beta_{j'm}^{IP,t} = 1$ is upheld. When the weight α is 0, and the corresponding weight β for undesirable outputs is 1, it signifies that the subsequent improvement strategies will prioritize the reduction of adverse outputs. This weighting strategy implies a dedication to minimizing undesirable outputs as much as possible without introducing any additional inputs, thereby enhancing industrial eco-efficiency. By meticulously adjusting these weights, we aspire to achieve substantial improvements in industrial eco-efficiency. It is noteworthy that the −0.00 in Table 3 is a positive number close to zero.

Table 3. Changes in Variables during the IP stage α from 2011 to 2021.

DMU	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Beijing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	−0.00	−0.00	0.915	1.000	−0.00	−0.00	1.000	−0.00	−0.00	1.000	1.000
Hebei	−0.00	−0.00	0.704	0.019	1.000	0.146	1.000	0.163	−0.00	1.000	−0.00
Liaoning	1.000	−0.00	0.050	−0.00	−0.00	0.088	0.140	−0.00	−0.00	1.000	−0.00
Shanghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Jiangsu	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	1.000	−0.00	−0.00	0.536	1.000
Zhejiang	1.000	1.000	−0.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Fujian	0.027	1.000	0.218	0.071	−0.00	1.000	0.267	−0.00	−0.00	0.306	0.204
Shandong	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Guangdong	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hainan	1.000	1.000	1.000	1.000	1.000	1.000	−0.00	−0.00	−0.00	−0.00	−0.00
Shanxi	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Jilin	−0.00	−0.00	1.000	−0.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Heilongjiang	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Anhui	1.000	−0.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.491	−0.00
Jiangxi	1.000	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
Henan	−0.00	−0.00	1.000	1.000	−0.00	−0.00	0.273	−0.00	0.142	0.329	0.241
Hubei	−0.00	0.143	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hunan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Inner Mongolia	0.336	−0.00	1.000	−0.00	1.000	0.179	−0.00	1.000	1.000	−0.00	0.209
Guangxi	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
Chongqing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sichuan	−0.00	1.000	0.001	0.246	−0.00	0.826	0.394	0.033	1.000	−0.00	1.000
Guizhou	0.512	1.000	0.176	−0.00	−0.00	1.000	1.000	−0.00	−0.00	1.000	0.324
Yunnan	1.000	1.000	0.221	0.125	−0.00	−0.00	1.000	0.049	−0.00	−0.00	1.000
Shaanxi	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	−0.00	1.000
Gansu	−0.00	−0.00	1.000	0.399	1.000	0.182	0.119	0.020	−0.00	0.357	1.000
Qinghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	−0.00	−0.00
Ningxia	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	0.600	1.000
Xinjiang	1.000	1.000	0.000	1.000	1.000	−0.00	0.457	−0.00	−0.00	0.165	−0.00

4.2. Comparisons of Original, Bootstrapping and Fuzzy Estimates

Figure 3 shows the original efficiency, bootstrap efficiency, and fuzzy efficiency, which are obtained through distinct methodologies for comprehensive analysis. The "original" represents the IPEGHT efficiency derived from the dynamic three-stage endogenous DDF model proposed in this paper, while the "bootstrap" refers to the IPEGHT efficiency obtained by incorporating the generalized smooth bootstrap method into the dynamic three-stage endogenous DDF model. Finally, the "fuzzy" refers to the integrated framework proposed in this paper, which combines the generalized smooth bootstrap method and fuzzy comparison methods within the dynamic three-stage endogenous DDF model.

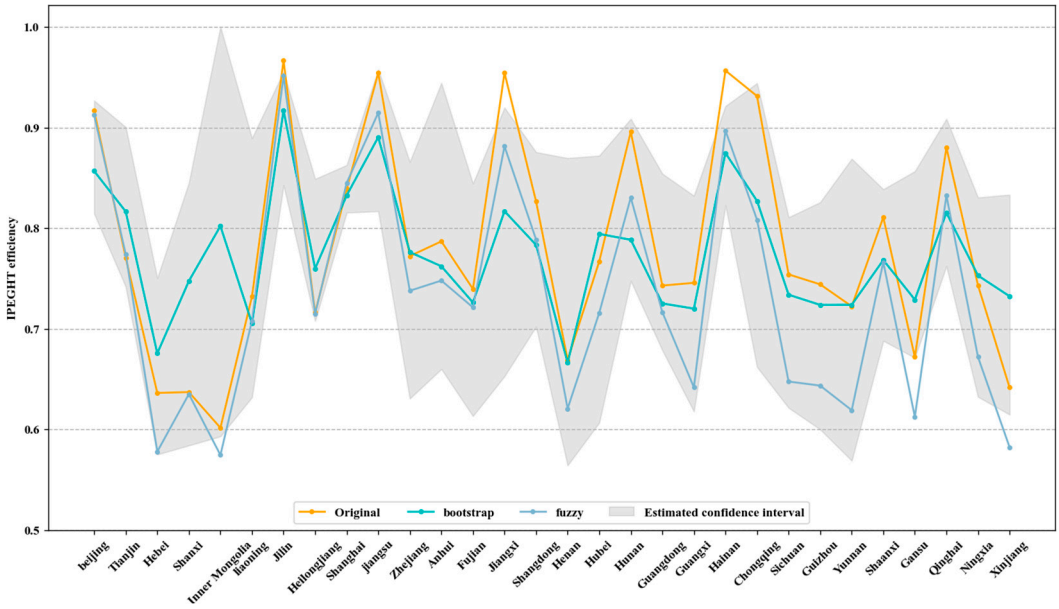


Figure 3. Comparison of DEA ecology system efficiency across 30 provinces using different methods.

The "original" line graph, which depicts the ecological efficiency of various provinces, is calculated using Equation (5). In this computation process, the input and output data employed are not subjected to fuzzification and regarded as precise numerical values. Furthermore, the "bootstrap" line graph is derived by introducing the bootstrap method based on the "original" scores, significantly enhancing the robustness of the DEA efficiency scores. Compared to the original scores, it is clearly visible that the bootstrap line exhibits certain fluctuations in some provinces, which intuitively reflects the effective deviation correction of the scores on the original ecological efficiency after applying the bootstrap method. It is noteworthy that before and after the bootstrap, the score of Inner Mongolia increased from 0.6 to 0.8. This paper adopts the generalized smooth bootstrap method to construct a 95% confidence interval, which is visually represented by a shaded area. After 2000 bootstrap iterations, all provinces fall within the 95% confidence interval.

The "Fuzzy" represents the process of fuzzifying precise data, calculating it using Equation (10) with fuzzy data, and applying the generalized fuzzy bootstrap method for bias correction, which more accurately reflects the actual situation of China's industry. This framework effectively handles missing values and uncertainties, correcting inherent biases in the DEA model. For example, in the case of Inner Mongolia shown in Figure 3, the presence of missing values meant that the bootstrap method alone did not adequately correct the bias. However, by integrating fuzzy comparison methods and the generalized bootstrap method, the results more accurately reflect the true situation.

From the Figure 3, it is evident that "original" represents traditional DEA, which generally yields higher overall efficiency due to the potential overestimation of IPEGHT efficiency. Zhuang et al. (2022) [6] employed the traditional DDF model to assess industrial efficiency, resulting in generally higher efficiency scores. However, the calculated efficiency results contain significant errors, such as the high efficiency reported for Yunnan, which does not align with the actual situation. The "bootstrap" appears smoother compared to the "original", indicating that the bootstrap method has a stabilizing effect on the original scores. This is likely due to the bootstrap method's ability to reduce random fluctuations through repeated sampling and averaging, enhance estimation accuracy by accounting for sample diversity and uncertainty, and handle outliers more robustly, thereby minimizing their impact on the overall results. Wan et al. (2023) [4] utilized the bootstrap method to optimize the traditional DEA model for assessing industrial efficiency across

Chinese provinces. The efficiency results obtained align closely with those derived from the generalized smoothing bootstrap method employed in this study. However, our approach involves fewer assumptions and is applicable to a broader range of scenarios.

The ecological efficiency calculated by the fuzzy DDF model exhibits significant fluctuations, with the efficiency of most provinces decreasing. However, the fuzzification process, which requires consideration of data uncertainty and external environmental changes such as data missingness and policy adjustments, once again led to increased volatility in scores. These factors may influence the efficiency scores after fuzzification and result in larger efficiency discrepancies among provinces. In some provinces, the differences between the three curves are quite significant, reflecting the pronounced impact of fuzzification on efficiency scores, which may be attributed to greater uncertainty in the data of these provinces. For example, provinces such as Yunnan, Gansu, and Xinjiang exhibit lower efficiency after fuzzification, which more accurately reflects the actual situation.

Table 4 presents the original efficiency obtained by applying the model proposed in this paper to the three stages of the IPEGHT eco-system. Additionally, the table compares the bootstrap efficiency derived from the generalized bootstrap method and the fuzzy efficiency obtained by incorporating fuzzy mathematical theory. A detailed comparison of the other three methods' efficiencies is provided, facilitating an in-depth analysis and understanding of the performance of each stage under different assessment methods.

Table 4. IPEGHT eco-system stage efficiency comparison under three assessment methods.

Region	DMU	IP			EG			HT		
		Original	**Bootstrap**	**Fuzzy**	**Original**	**Bootstrap**	**Fuzzy**	**Original**	**Bootstrap**	**Fuzzy**
Eastern	Beijing	1	1	1	0.991	0.977	0.971	0.799	0.659	0.803
	Tianjin	0.759	0.842	0.738	0.476	0.548	0.516	1	0.999	0.995
	Hebei	0.397	0.476	0.313	0.423	0.472	0.435	0.976	0.978	0.884
	Liaoning	0.577	0.581	0.551	0.53	0.471	0.5	1	0.974	0.982
	Shanghai	1	1	1	0.483	0.49	0.507	0.985	0.964	0.982
	Jiangsu	1	1	1	0.899	0.875	0.866	0.962	0.82	0.888
	Zhejiang	0.907	0.906	0.824	0.563	0.603	0.592	0.828	0.809	0.782
	Fujian	0.985	0.982	0.918	0.232	0.281	0.376	0.936	0.868	0.833
	Shandong	0.845	0.882	0.862	0.671	0.666	0.661	0.929	0.797	0.829
	Guangdong	1	1	1	0.279	0.289	0.312	0.898	0.846	0.808
	Hainan	1	1	1	0.994	0.978	0.941	0.896	0.703	0.787
Eastern mean		0.861	0.879	0.837	0.595	0.605	0.607	0.928	0.856	0.87
Central	Shanxi	0.567	0.758	0.577	0.239	0.412	0.287	0.988	0.991	0.94
	Jilin	1	0.998	0.998	0.892	0.855	0.891	0.998	0.903	0.962
	Heilongjiang	0.586	0.654	0.569	0.468	0.545	0.48	1	1	1
	Anhui	0.894	0.841	0.817	0.396	0.444	0.423	1	0.942	0.94
	Jiangxi	1	0.884	0.984	0.848	0.701	0.713	1	0.853	0.93
	Henan	0.69	0.724	0.613	0.215	0.244	0.293	0.994	0.94	0.872
	Hubei	0.893	0.948	0.827	0.43	0.526	0.439	0.925	0.88	0.839
	Hunan	0.728	0.617	0.599	0.945	0.928	0.937	0.985	0.812	0.924
Central mean		0.795	0.803	0.748	0.554	0.582	0.558	0.986	0.915	0.926
Western	Inner Mongolia	0.681	0.87	0.632	0.207	0.567	0.234	0.838	0.928	0.787
	Guangxi	0.611	0.63	0.571	0.541	0.503	0.356	1	0.951	0.909
	Chongqing	0.945	0.816	0.78	0.835	0.752	0.663	0.994	0.892	0.939
	Sichuan	0.754	0.755	0.623	0.457	0.46	0.333	0.977	0.923	0.902
	Guizhou	0.721	0.723	0.559	0.445	0.445	0.36	0.986	0.934	0.92
	Yunnan	0.704	0.751	0.585	0.447	0.497	0.402	0.943	0.873	0.808
	Shaanxi	0.952	0.938	0.82	0.551	0.531	0.566	0.9	0.819	0.876
	Gansu	0.584	0.716	0.495	0.361	0.44	0.391	0.971	0.955	0.866
	Qinghai	0.992	0.971	0.918	0.785	0.796	0.757	0.868	0.713	0.824
	Ningxia	0.974	0.989	0.916	0.385	0.419	0.297	0.839	0.826	0.77
	Xinjiang	0.851	0.929	0.749	0.282	0.416	0.29	0.755	0.822	0.677
Western mean		0.797	0.826	0.695	0.481	0.53	0.423	0.915	0.876	0.843
National mean		0.82	0.839	0.761	0.542	0.571	0.527	0.939	0.879	0.876

4.3. Analysis of the Fuzzy Efficiency

Figure 4 analyzes the impact of carry-over factors on the efficiency of the IPEGHT eco-system. The dynamic analysis that considers carry-over factors reveals long-term efficiency issues, whereas the static analysis without carry-over focuses only on the efficiency of a single time point. In this study, carry-over refers to industrial fixed assets, which may not have been fully utilized at a given time point but are treated as fully utilized in the calculations, leading to an overestimation of efficiency. The efficiency of industrial fixed assets in each province is continuous, with investments from the previous year affecting subsequent years. Dynamic analysis requires considering resource allocation for both the current and previous years, increasing model constraints and potentially leading to lower efficiency scores. In actual production, resources are intertemporal, and models considering carry-over factors more accurately reflect the dynamic utilization of resources.

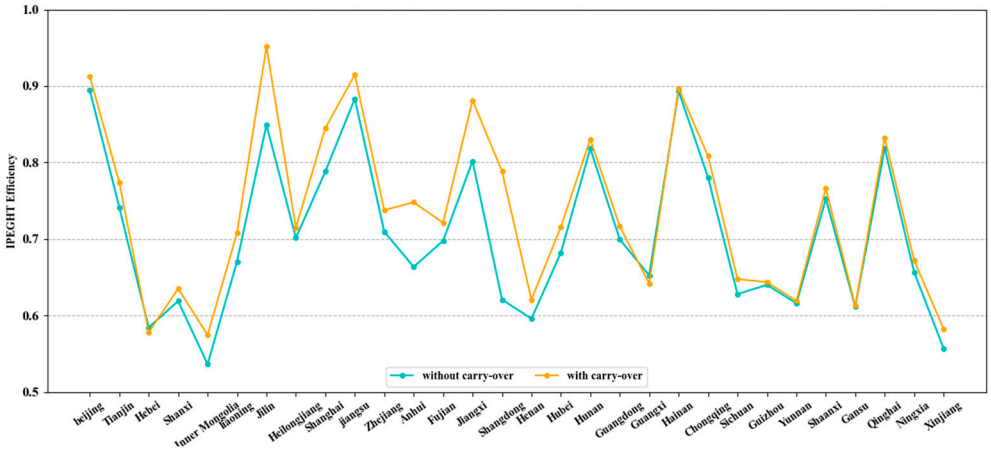


Figure 4. Analysis of the impact of carry-over factors on IPEGHT eco-system efficiency.

From Table 5, it is evident that the ecological efficiency scores of most provinces exhibited fluctuations across different years. Specifically, certain provinces such as Beijing, Jiangsu, and Jilin have achieved notably high scores in some years, demonstrating a high level of ecological efficiency. In contrast, provinces like Hebei, Shanxi, Qinghai, and Ningxia have relatively lower scores, indicating a lack of balance between industrial development, environmental impact, and health protection. There is a pressing need for significant improvement in these regions.

From a regional disparity perspective, the eastern coastal developed provinces, such as Beijing, Shanghai, Jiangsu, and Zhejiang, generally achieved higher scores in most years, highlighting their significant advantages in environmental protection, resource utilization efficiency, and health protection. In contrast, the western and central regions, including Gansu, Qinghai, Ningxia, and Xinjiang, consistently achieved lower scores, suggesting that there is still room to improve in these areas. From a temporal trend perspective, some provinces such as Beijing, Jiangsu, and Zhejiang have shown either an overall upward trend or maintained high scores, indicating that these provinces may be progressively enhancing their ecological efficiency. However, there are also provinces with significant score fluctuations, such as Fujian and Liaoning, which indicates that these provinces have undergone substantial adjustments or are facing challenges in terms of ecological efficiency. It is noteworthy that, in certain years, the scores of specific provinces experienced significant abnormal fluctuations. For instance, Liaoning showed a substantial decline in its score in 2018. As an old industrial base, Liaoning has a significant proportion of traditional heavy industries. With the advancement of national industrial structure adjustment and the implementation of transformation and upgrading policies, Liaoning is actively optimizing

its industrial structure, shifting from heavy industries to high-tech industries. The decline of old industries and the rise of new ones may lead to efficiency fluctuations.

Table 5. Changes in IPEGHT eco-system efficiency across 30 provinces in China, 2011–2021.

Years DMUs	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Overall
Beijing	1	0.886	0.901	0.861	0.885	0.93	0.912	0.941	1	0.889	0.964	0.912
Tianjin	0.649	0.824	0.83	0.617	0.71	0.727	0.802	0.551	0.673	0.931	0.932	0.774
Hebei	0.437	0.638	0.417	0.443	0.761	0.652	0.776	0.39	0.377	0.608	0.484	0.578
Liaoning	0.933	0.62	0.739	0.562	0.783	0.559	0.815	0.552	0.46	0.806	0.625	0.708
Shanghai	0.791	1	0.995	0.706	1	1	0.869	0.675	0.693	0.697	0.702	0.845
Jiangsu	0.845	1	0.994	0.977	0.838	0.978	0.968	0.7	0.954	0.927	0.914	0.915
Zhejiang	0.876	0.847	0.928	0.896	0.623	0.992	0.67	0.61	0.468	0.635	0.517	0.738
Fujian	0.54	0.793	0.766	0.636	0.544	0.856	0.816	0.677	0.772	0.747	0.651	0.721
Shandong	0.506	0.776	0.676	0.964	0.937	0.919	0.954	0.919	0.918	0.41	0.649	0.789
Guangdong	0.626	0.793	0.731	0.74	0.811	0.699	0.561	0.628	0.827	0.691	0.664	0.717
Hainan	1	0.908	0.897	0.897	0.902	0.872	0.795	0.964	0.929	0.952	0.886	0.897
Shanxi	0.578	0.544	0.596	0.532	0.505	0.702	0.589	0.441	0.527	0.889	0.709	0.635
Jilin	0.889	1	0.985	1	0.985	0.904	0.978	0.999	0.762	1	0.952	0.952
Heilongjiang	0.637	0.809	0.78	0.54	0.667	0.892	0.799	0.552	0.657	0.529	0.652	0.715
Anhui	0.552	0.538	0.781	0.861	0.782	1	0.86	0.748	0.633	0.719	0.521	0.748
Jiangxi	0.827	0.971	0.998	0.994	0.877	0.841	0.967	0.739	0.646	0.839	0.934	0.881
Henan	0.649	0.61	0.688	0.821	0.522	0.515	0.53	0.524	0.453	0.56	0.648	0.621
Hubei	0.576	0.662	0.958	0.962	0.746	0.606	0.692	0.756	0.676	0.533	0.556	0.716
Hunan	0.754	0.698	0.847	0.885	0.916	0.844	0.82	0.9	0.824	0.78	0.754	0.83
Inner Mongolia	0.591	0.644	0.428	0.564	0.694	0.463	0.469	0.424	0.42	0.457	0.905	0.575
Guangxi	0.504	0.588	0.534	0.678	0.581	0.583	0.605	0.777	0.93	0.551	0.402	0.642
Chongqing	0.779	0.731	0.803	0.892	0.889	0.869	0.971	0.809	0.504	0.649	0.837	0.808
Sichuan	0.407	0.513	0.637	0.636	0.569	0.542	0.768	0.608	0.533	0.599	1	0.648
Guizhou	0.534	0.543	0.559	0.768	0.719	0.596	0.854	0.566	0.407	0.622	0.576	0.644
Yunnan	0.433	0.9	0.451	0.471	0.807	0.449	0.754	0.563	0.442	0.601	0.709	0.619
Shaanxi	0.713	0.827	0.58	0.854	0.879	0.615	0.982	0.823	0.762	0.405	0.855	0.766
Gansu	0.434	0.52	0.452	0.56	0.854	0.634	0.644	0.546	0.472	0.499	0.812	0.612
Qinghai	0.484	0.906	0.903	0.979	0.911	0.812	0.905	0.856	0.691	0.821	0.899	0.833
Ningxia	0.647	0.839	0.594	0.585	0.785	0.685	0.636	0.712	0.646	0.599	0.543	0.672
Xinjiang	0.378	0.632	0.613	0.466	0.722	0.831	0.533	0.498	0.435	0.568	0.614	0.582

From 2011 to 2021, the average score of fuzzy efficiency reached 0.736, which amply demonstrates that the overall performance of China's IPEGHT eco-system is quite commendable. Notably, in terms of IPEGHT fuzzy efficiency, only Inner Mongolia and Xinjiang scored 0.575 and 0.582 respectively, while the IPEGHT efficiency of the other 28 provinces exceeded 0.6. Figure 5 further unveils the detailed efficiency results of the IPEGHT eco-system.

Focusing on the IP stage, it can be observed that from 2011 to 2021, the efficiency of Beijing, Shanghai, Jiangsu, Guangdong, and Hainan consistently remained at the perfect score of 1. Simultaneously, Jilin and Jiangxi also maintained the same efficiency level. This data comparison reveals that the overall efficiency of the eastern region is significantly superior to that of the central and western regions during the IP stage. This conclusion aligns with the findings of Wan et al. (2023) [4]. During the EG stage, the overall efficiency scores were relatively low. Specifically, Beijing maintained a perfect score of 1 in all years except 2021, while Hainan and Jiangsu also scored 1 for most of the time. Hunan similarly scored 1 for most of the time, whereas the western region generally had lower scores for most of the time. In the HT stage, the overall efficiency significantly improved, and the gap between regions notably narrowed. This phenomenon clearly demonstrates that the government has placed a high priority on protecting the health of its people and achieved remarkable results.

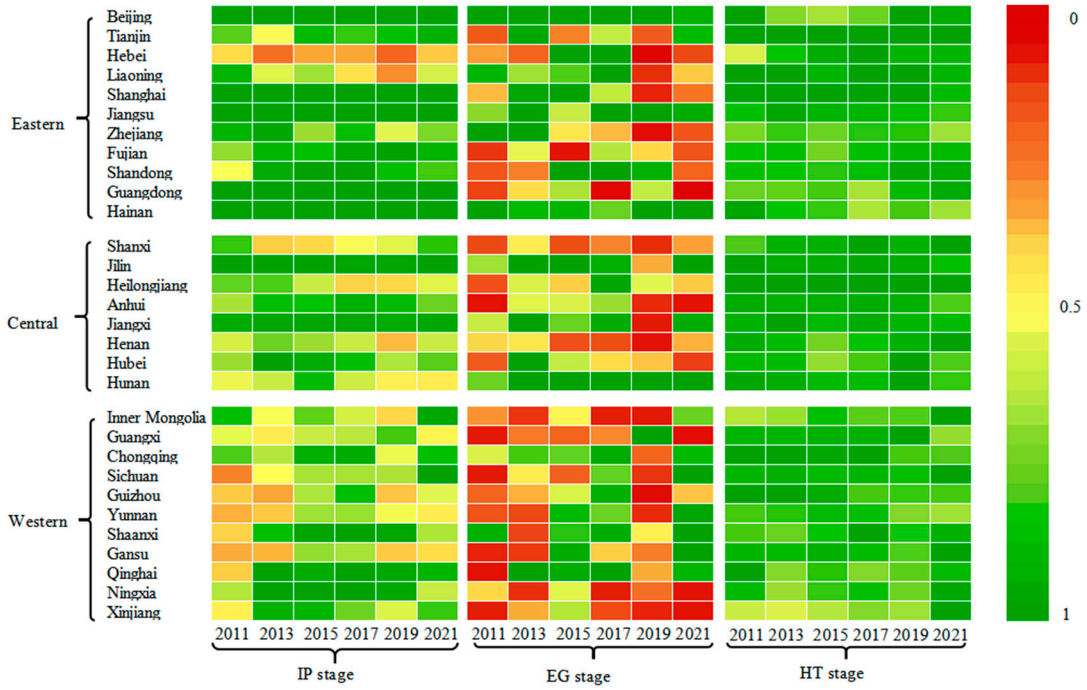


Figure 5. Detailed Efficiency Analysis for 30 Provinces from 2011 to 2021.

Figure 6 consists of four kernel density estimation plots, illustrating the efficiency changes of the IPEGHT eco-system and its stages. To delve into the temporal and spatial variations of IPEGHT industrial ecological efficiency, and to uncover the underlying reasons behind these variations, the paper employs the kernel density estimation method. It can be observed that the overall efficiency of IPEGHT exhibits a relatively dispersed score distribution, and the fluctuations have a significant impact on the overall efficiency during the EG stage. Despite variations among provinces in the efficiency of the overall eco-system and its stages, the efficiency in IP stage and HT stage was stable and high. In contrast, the efficiency of EG was more volatile, with greater uncertainty and significant differences among provinces. It reflects the differential impacts of factors such as policy implementation, resource investment, and technological advancements in environmental protection across different years. China has been striving to balance industrial production, environmental protection, and health protection over these 11 years, but there is still room to enhance overall system efficiency, particularly in the realm of environmental protection.

In the IP stage, efficiency scores predominantly clustered within the higher value range, particularly in the vicinity of the 1.0 high-score bracket. This phenomenon indicates that the IP efficiency scores of most provinces were generally high during this period. Throughout the entire research timeframe, the fluctuations in score distribution were relatively limited, with the efficiency scores of most provinces exhibiting a stable trend. Nevertheless, in certain individual years, there was a slight increase in the number of provinces with low efficiency, reflecting some fluctuations in efficiency levels in localized areas. With continuous technological advancements and the introduction of policies supporting industrial production by the government, the efficiency of IP stage has significantly been improved, thereby maintaining high efficiency scores for most provinces. However, policy adjustments may negatively impact the industrial production efficiency of certain provinces.

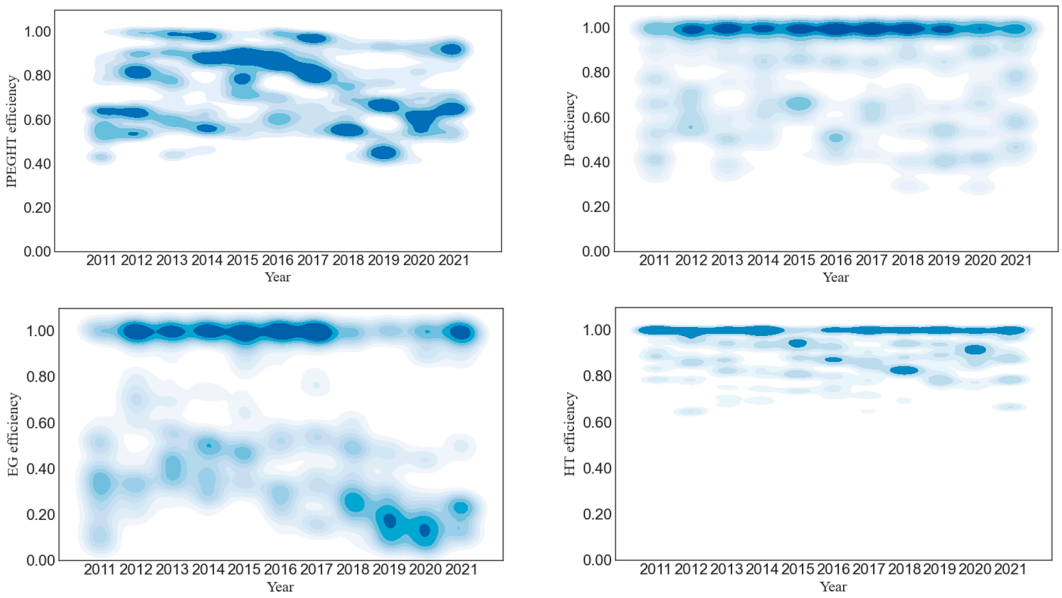


Figure 6. The changing trends of IPEGHT eco-system efficiency and its stage efficiencies.

In the EG stage, the distribution of efficiency scores is relatively dispersed, indicating significant differences among provinces. Particularly in 2019 and 2020, the ecological efficiency scores of multiple provinces were relatively low. However, this situation showed a marked improvement in 2021. The EG efficiency scores exhibit multiple peaks, reflecting a complex distribution of efficiency scores across provinces during this stage, rather than a concentration around a specific value. Certain provinces face greater governance challenges due to high pollution emissions, and the varying degrees of policy implementation across provinces lead to dispersed efficiency scores. Provinces with higher economic development levels are better equipped to allocate resources for environmental governance.

In the HT stage, efficiency has remained relatively stable, with scores in most years concentrated between 0.6 and 1.0, indicating relatively effective health protection management. The peak in the high-score range suggests that over time, an increasing number of provinces have performed well in HT efficiency, gradually approaching an optimal state. The advancement of medical technology and the heightened awareness of public health have prompted provinces to place greater emphasis on health protection, leading to increased resource allocation for health protection.

Figure 7 delves into the efficiency scores of the IPEGHT eco-system and its stages across the eastern, central, and western regions, revealing the dynamic trends in efficiency within each region from 2011 to 2021. For the overall IPEGHT eco-system, the western region initially lagged the eastern and central regions in terms of efficiency, starting from 0.5 in 2011 and fluctuating upwards to reach 0.75 by 2021. The eastern region, on the other hand, consistently maintained a high level of efficiency throughout most of the period, exhibiting minimal fluctuations. In contrast, the central region experienced more pronounced fluctuations in its efficiency scores. In the IP stage, the eastern region showed extremely stable performance, and the scoring efficiency has continued to remain at a high level. In contrast, the western region showed unusually severe fluctuations and relatively low efficiency. In the EG stage, fluctuations in the eastern, central, and western regions were particularly severe, and efficiency scores were generally low, with scores ranging from 0.2 to 0.8. This phenomenon indicates that all provinces still need to attach great importance to the treatment of industrial "three wastes" and intensify efforts to improve and optimize it. In HT stage, the efficiency scores were notably high, consistently exceeding 0.8, and exhibited remarkable stability.

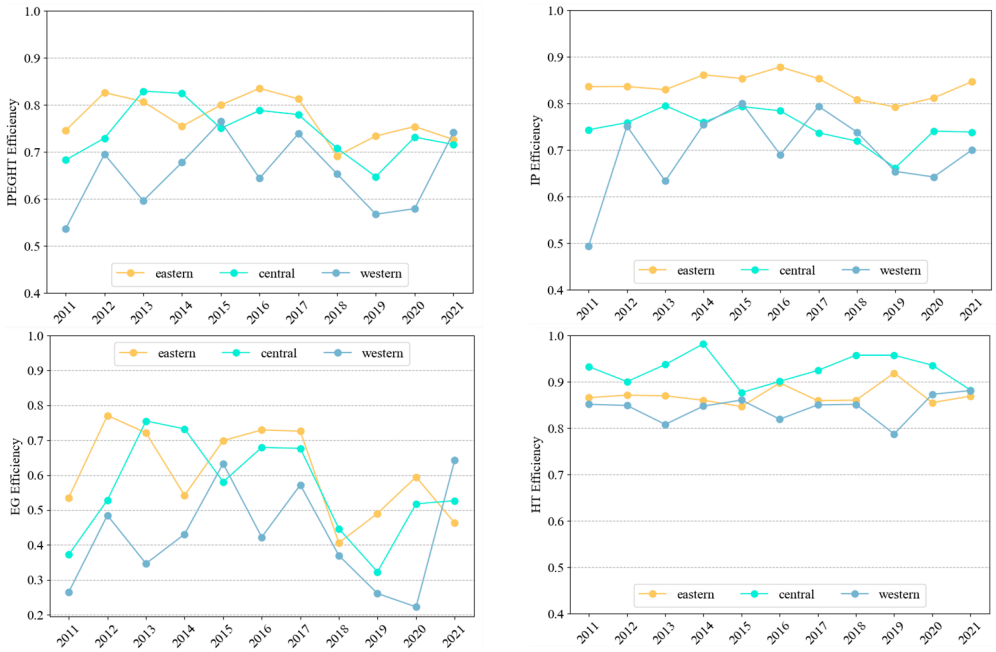


Figure 7. Dynamic changes of the IPEGHT system and its stages across regions.

5. Conclusions

A dynamic three-stage DDF model is proposed in this paper, and an endogenous direction vector selection method is introduced. The generalized bootstrap method is adopted with an integrated approach to correct the inherent bias issues in traditional DEA methods. Subsequently, fuzzy mathematics is employed to handle data uncertainties in the real world. This model is employed to evaluate the efficiency of the IPEGHT industrial eco-system. The main conclusions include:

Firstly, the model proposed in this paper enables it to handle data uncertainty and enhance the robustness of the model. This approach more accurately reflects the industrial ecological efficiency of provinces. By employing the generalized smooth bootstrap method, we have effectively addressed the inherent bias in the original estimation. After conducting 2000 bootstrap iterations, the dynamic three-stage DDF model demonstrates significant stability at a 95% confidence level, despite some fluctuations in bootstrap bias. In response to the uncertainty issues inherent in real-world data, this paper introduces the fuzzy mathematics theory and utilizes the LU-ordering method to compare fuzzy trapezoidal numbers.

Secondly, the paper introduces a straightforward method for selecting endogenous direction vectors, effectively guiding each province in enhancing its efficiency and setting itself apart from other approaches that require complex mathematical models and algorithms for the same purpose. The proposed method simplifies the selection of the endogeneity direction vector by determining the weight vector based on the data of the DMUs themselves, thereby offering a clear direction for enhancing industrial ecological efficiency across various provinces.

Thirdly, there is still significant room for improvement in China's industrial eco-efficiency, particularly during the EG stage. By employing the generalized smoothing bootstrap method to correct the raw efficiency scores, we observed that the IPEGHT efficiency exhibited a smoother trend compared to the original efficiency. However, upon fuzzifying the data, fluctuations emerged in the efficiency curve. Notably, after fuzzification, the eco-efficiency scores declined, whereas the original eco-efficiency scores remained relatively high.

Finally, it is noteworthy that apart from the HT stage, the eastern region significantly outperforms the central and western regions in both overall and stage-wise performance, with relatively stable efficiency. Meanwhile, the central region demonstrates a notable momentum of progress. The western region experiences significant fluctuations in efficiency during the IP and EG stages, with low efficiency levels, and registers the lowest efficiency during the HT stage.

This model supports the achievement of the United Nations Sustainable Development Goals, playing a significant role in the fields of industrial production, environmental governance, and health threats. For specific targets in different regions of China, the eastern region needs to maintain high efficiency while continuously optimizing environmental policies; the central region, while making progress, should increase investment in environmental governance to narrow the gap with the eastern region; and the western region should focus on improving the efficiency of industrial production and environmental governance to reduce health threats. This provides clear directions for sustainable development in each region.

The dynamic three-stage DDF model and its framework proposed in this study not only offer a new perspective for assessing industrial eco-efficiency but also hold profound implications for multiple industries in practical applications. Specifically, the model provides a quantitative tool for policymakers and business managers to identify and optimize resource allocation, thereby enhancing eco-efficiency and reducing environmental risks. In the energy sector, the model can be utilized to evaluate and improve energy utilization efficiency, driving the development and adoption of renewable energy technologies. The environmental protection industry will benefit from more precise assessments of environmental governance measures, enhancing the effectiveness of pollution control and ecological restoration projects. In the public health domain, the model can be employed to evaluate the impact of environmental policies on public health, optimize resource allocation, and improve the capacity for disease prevention and health promotion. Additionally, policy-making departments can leverage the model's evaluation results to formulate or adjust relevant policies, promoting sustainable development of the economy, environment, and society. In the long run, the application of this model will foster cross-industry collaboration, stimulate innovation in green technologies and sustainable practices, and provide support for achieving global sustainable development goals.

Future research will advance parameter optimization and data fuzzification in fuzzy DEA models, aiming to boost practical applicability. Integration with machine learning and deep learning will be explored to forecast efficiency trends, simulate decisions, and provide precise, personalized optimization solutions using big data and complex models.

Author Contributions: Conceptualization, X.W. and Q.P.; methodology, X.W., Q.P. and L.Z.; software, X.W.; validation, Q.P. and L.Z.; formal analysis, X.W.; investigation, X.W.; resources, X.W.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, Q.P., L.Z. and Y.S.; visualization, X.W.; supervision, Q.P., L.Z. and Y.S.; project administration, Q.P.; funding acquisition, Q.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [the Humanities and Social Science Research General Project of the Ministry of Education of China] grant number [No. 22YJAZH086] and [Foundation of Philosophy and Social Science for Universities in Jiangsu Province] grant number [No. 2023SJYB1235]. And The APC was funded by [No. 22YJAZH086].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data availability is restricted to protect participant privacy.

Acknowledgments: All individuals mentioned in the acknowledgements section have provided their consent to be acknowledged.

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

References

1. Yang, L.; Qian, N.; Niu, Y.; Lu, J. Dynamic Association between Industrialand Agricultural Economic Development, Environmental Pollution and PublicHealth in China: Based on the ParallelTwo-Stage EBM-DEA Model. *Pol. J. Environ. Stud.* **2024**, *33*, 3443–3461. [CrossRef] [PubMed]
2. Tone, K.; Tsutsui, M. Dynamic DEA with network structure: A slacks-based measure approach. *Omega* **2014**, *42*, 124–131. [CrossRef]
3. Song, M.; Wang, J. Environmental efficiency evaluation of thermal power generation in China based on a slack-based endogenous directional distance function model. *Energy* **2018**, *161*, 325–336. [CrossRef]
4. Wan, L.; Zhang, L.; Chiu, Y.; Pang, Q.; Luo, Q.; Shi, Z. A bootstrapping dynamic two-stage SBM model: An application to assess industrial water use and health risk systems. *Sci. Total Environ.* **2023**, *894*, 164813. [CrossRef] [PubMed]
5. Zhou, W.; Xu, Z. An Overview of the Fuzzy Data Envelopment Analysis Research and Its Successful Applications. *Int. J. Fuzzy Syst.* **2020**, *22*, 1037–1055. [CrossRef]
6. Zhuang, Y.; Zhang, L.; Chiu, Y.; Guo, Z.; Chu, Y.; Du, X. Evaluating the recycling efficiency of industrial water use systems in China: Basin differences and factor analysis. *J. Environ. Manag.* **2022**, *316*, 115313. [CrossRef]
7. Simar, L.; Wilson, P.W. Statistical Approaches for Non-parametric Frontier Models: A Guided Tour. *Int. Stat. Rev.* **2015**, *83*, 77–110. [CrossRef]
8. Simar, L.; Wilson, P.W. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Manag. Sci.* **1998**, *44*, 49–61. [CrossRef]
9. Simar, L.; Wilson, P.W. A general methodology for bootstrapping in non-parametric frontier models. *J. Appl. Stat.* **2000**, *27*, 779–802. [CrossRef]
10. Kneip, A.; Simar, L.; Wilson, P.W. Asymptotics and Consistent Bootstraps for DEA Estimators in Nonparametric Frontier Models. *Econom. Theory* **2008**, *24*, 1663–1697. [CrossRef]
11. Simar, L.; Vanhems, A.; Wilson, P.W. Statistical inference for DEA estimators of directional distances. *Eur. J. Oper. Res.* **2012**, *220*, 853–864. [CrossRef]
12. Simar, L.; Wilson, P.W. Inference by the m out of n bootstrap in nonparametric frontier models. *J. Product. Anal.* **2011**, *36*, 33–53. [CrossRef]
13. Lin, S.-W.; Lu, W.-M. A comparison of chance-constrained data envelopment analysis, stochastic nonparametric envelopment of data and bootstrap method: A case study of cultural regeneration performance of cities. *Eur. J. Oper. Res.* **2024**, *316*, 1179–1191. [CrossRef]
14. Aparicio, J.; Kapelko, M.; Ortiz, L. Enhancing the measurement of firm inefficiency accounting for corporate social responsibility: A dynamic data envelopment analysis fuzzy approach. *Eur. J. Oper. Res.* **2023**, *306*, 986–997. [CrossRef]
15. Kao, C. Fuzzy efficiency measures in data envelopment analysis. *Fuzzy Sets Syst.* **2000**, *113*, 427–437. [CrossRef]
16. Arana-Jiménez, M.; Sánchez-Gil, M.C.; Lozano, S. A fuzzy DEA slacks-based approach. *J. Comput. Appl. Math.* **2022**, *404*, 113180. [CrossRef]
17. Mozaffari, M.R.; Mohammadi, S.; Wanke, P.F.; Correa, H.L. Towards greener petrochemical production: Two-stage network data envelopment analysis in a fully fuzzy environment in the presence of undesirable outputs. *Expert Syst. Appl.* **2021**, *164*, 113903. [CrossRef]
18. Zhang, S.; Li, S.; Zhang, S.; Zhang, M. Decision of Lead-Time Compression and Stable Operation of Supply Chain. *Complexity* **2017**, *2017*, 7436764. [CrossRef]
19. Huang, B.; Miao, J.; Li, Q. A Vetoed Multi-objective Grey Target Decision Model with Application in Supplier Choice. *J. Grey Syst.* **2022**, *34*, 15.
20. Li, C.; He, A.; Liu, G.; Wen, Y.; Chronopoulos, A.T.; Giannakos, A. RFL-APIA: A Comprehensive Framework for Mitigating Poisoning Attacks and Promoting Model Aggregation in IIoT Federated Learning. *IEEE Trans. Ind. Inform.* **2024**, *20*, 12935–12944. [CrossRef]
21. Hatami-Marbini, A.; Emrouznejad, A.; Tavana, M. A taxonomy and review of the fuzzy data envelopment analysis literature: Two decades in the making. *Eur. J. Oper. Res.* **2011**, *214*, 457–472. [CrossRef]
22. Wu, H.-C. The optimality conditions for optimization problems with convex constraints and multiple fuzzy-valued objective functions. *Fuzzy Optim. Decis. Mak.* **2009**, *8*, 295–321. [CrossRef]
23. Stefanini, L.; Arana-Jiménez, M. Karush–Kuhn–Tucker conditions for interval and fuzzy optimization in several variables under total and directional generalized differentiability. *Fuzzy Sets Syst.* **2019**, *362*, 1–34. [CrossRef]

24. Tone, K.; Tsutsui, M. Dynamic DEA: A slacks-based measure approach. *Omega* **2010**, *38*, 145–156. [CrossRef]
25. Färe, R.; Grosskopf, S.; Whittaker, G. Directional output distance functions: Endogenous directions based on exogenous normalization constraints. *J. Product. Anal.* **2013**, *40*, 267–269. [CrossRef]
26. Fukuyama, H.; Weber, W.L. Measuring Japanese bank performance: A dynamic network DEA approach. *J. Product. Anal.* **2015**, *44*, 249–264. [CrossRef]
27. Fukuyama, H.; Weber, W.L. Japanese Bank Productivity, 2007–2012: A Dynamic Network Approach. *Pac. Econ. Rev.* **2017**, *22*, 649–676. [CrossRef]
28. Fukuyama, H.; Tsionas, M.; Tan, Y. Dynamic network data envelopment analysis with a sequential structure and behavioural-causal analysis: Application to the Chinese banking industry. *Eur. J. Oper. Res.* **2023**, *307*, 1360–1373. [CrossRef]
29. Sueyoshi, T.; Yuan, Y.; Goto, M. A literature study for DEA applied to energy and environment. *Energy Econ.* **2017**, *62*, 104–124. [CrossRef]
30. Zhou, Y.; Kong, Y.; Zhang, T. The spatial and temporal evolution of provincial eco-efficiency in China based on SBM modified three-stage data envelopment analysis. *Environ. Sci. Pollut. Res.* **2020**, *27*, 8557–8569. [CrossRef]
31. Zhang, R.-L.; Liu, X.-H. Evaluating ecological efficiency of Chinese industrial enterprise. *Renew. Energy* **2021**, *178*, 679–691. [CrossRef]
32. Diao, Y.; Zhang, Q. Optimization of Management Mode of Small- and Medium-Sized Enterprises Based on Decision Tree Model. *J. Math.* **2021**, *2021*, 2815086. [CrossRef]
33. Wang, Q.; Tang, J.; Choi, G. A two-stage eco-efficiency evaluation of China's industrial sectors: A dynamic network data envelopment analysis (DNDEA) approach. *Process Saf. Environ. Prot.* **2021**, *148*, 879–892. [CrossRef]
34. Chen, Y.; Li, Q.; Liu, J. Innovating Sustainability: VQA-Based AI for Carbon Neutrality Challenges. *J. Organ. End User Comput.* **2024**, *36*, 1–22. [CrossRef]
35. Yang, F.; Wei, F.; Li, Y.; Huang, Y.; Chen, Y. Expected efficiency based on directional distance function in data envelopment analysis. *Comput. Ind. Eng.* **2018**, *125*, 33–45. [CrossRef]
36. Lozano, S.; Soltani, N. DEA target setting using lexicographic and endogenous directional distance function approaches. *J. Product. Anal.* **2018**, *50*, 55–70. [CrossRef]
37. Hampf, B.; Krüger, J.J. Optimal Directions for Directional Distance Functions: An Exploration of Potential Reductions of Greenhouse Gases. *Am. J. Agric. Econ.* **2015**, *97*, 920–938. [CrossRef]
38. Shao, L.; Yu, X.; Feng, C. Evaluating the eco-efficiency of China's industrial sectors: A two-stage network data envelopment analysis. *J. Environ. Manag.* **2019**, *247*, 551–560. [CrossRef]
39. Shabani, M.; Kordrostami, S.; Jahani Sayyad Noveiri, M. Renewable energy performance analysis using fuzzy dynamic directional distance function model under natural and managerial disposability. *Appl. Energy* **2023**, *352*, 121940. [CrossRef]
40. Wu, S.; Cheng, P.; Yang, F. Study on the impact of digital transformation on green competitive advantage: The role of green innovation and government regulation. *PLoS ONE* **2024**, *19*, e0306603. [CrossRef] [PubMed]
41. Nadarajan, D.; Ahmed, S.A.M.; Noor, N.F.M. Seaport Network Efficiency Measurement Using Triangular and Trapezoidal Fuzzy Data Envelopment Analyses with Liner Shipping Connectivity Index Output. *Mathematics* **2023**, *11*, 1454. [CrossRef]
42. Gerami, J.; Mozaffari, M.R.; Wanke, P.F.; Tan, Y. Fully fuzzy DEA: A novel additive slacks-based measure model. *Soft Comput.* **2023**. [CrossRef]
43. Nitta, T. Solving the XOR problem and the detection of symmetry using a single complex-valued neuron. *Neural Netw* **2003**, *16*, 1101–1105. [CrossRef] [PubMed]
44. NBSCES. *China Environmental Statistics Yearbook (2011–2022)*; China Statistical Press: Beijing, China, 2022.
45. NHC. *China Health Statistical Yearbook (2011–2022)*; Associated Medical University of China Press: Beijing, China, 2022.
46. NBSC. *China Statistical Yearbook (2011–2022)*; China Statistical Press: Beijing, China, 2022.
47. NBSPES. *China Population & Employment Statistical Yearbook (2011–2021)*; China Statistical Press: Beijing, China, 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.