Proceedings of Business and Economic Studies, 2026, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/PBES
Online ISSN: 2209-265X
Print ISSN: 2209-2641

Study on the Relationship between Green Development and Enterprise Profitability: An Empirical Analysis Based on Micro-Data of Chinese Industrial Enterprises

Junlin Chen

Nanjing University of Information Science and Technology, Jiangsu 210000, Nanjing, China

Abstract: Currently, the first problem that has gradually emerged from the development of modern industries is how to achieve ecological civilisation. With the ambitious goals of China's "Dual Carbon", we need to know how financial performance in enterprises relates to green development. Based on the data from 2018 to 2022, this paper will examine the relationship between the financial health, scale, and extent of environmental pollution of industrial enterprises in China. A reproducible Python/KNIME pipeline has been employed to process the data and train a model. Exploratory Data Analysis shows that the distribution of wastewater discharge intensity is right-skewed, and a small number of enterprises are responsible for a large amount of pollution. Based on the empirical model, there are several reasons for the irregular distribution of industrial pollution. The small R-squared of OLS is about 0.02, and Random Forest has reduced the average baseline error (MAE 1.25) by 74%. To avoid circular reasoning, K-means clustering was applied only to the financial and scale attributes, and the Silhouette score was also obtained. Post-hoc environmental analysis of the three derived financial clusters (Large-Scale Heavyweights, Mid-Sized Manufacturers and Low-Impact Micro-Enterprises) showed different structural emission patterns. Therefore, various environmental regulations and specific green finance will be implemented.

Keywords: Green development; Enterprise profitability; Environmental regulation; Random forest; Cluster analysis; Dual carbon goals

Online publication: June 19, 2026

1. Introduction

Amidst changes in the environment around the world, "green development" has become one of the main goals for all parts of the economy. In recent years, rapid industrialisation has been the main driving force for the economy and, in the process, harm to the environment. With the spread of awareness about climate change, many countries have shifted to sustainable development. At home, China needs to improve its economy and meet the demand for clean development simultaneously. The strategy to achieve the goals of "Dual Carbon", that is, to reduce carbon dioxide emissions to a peak before 2030 and achieve carbon neutrality by 2060, has imposed new regulations on the industrial sector [1]. Therefore, we need to know how small and medium-sized enterprises are responding to this change, whether they will be benefited by the increase in profits or whether, at the same time, the demand for environmental protection will also stimulate development [2].

The dataset is the detailed information of enterprises in China from the Industrial and Environmental Statistics (N = 45,218 valid firm-level observations from 2018 to 2022). It includes many attributes of the company, such as basic information, economic and financial indicators, environmental pollution indicators, etc. Many small-scale data can be used to better understand the variety of enterprises.

To help verify the results of the analysis, both Python is used for optimization of the algorithm and KNIME is employed to display a clear workflow. First, it will be determined how the firm's profitability, physical size and the extent of environmental damage are related. Using microdata, we can examine the heterogeneity of firms; it can be found that linear models are not suitable for the base case comparison; and derive a non-circular financial classification to quantify environmental burden.

2.Literature review

The neoclassical view treats environmental regulation as a deadweight loss, whereas the Porter Hypothesis suggests welldesigned rules can induce "innovation offsets" [3]. Under China's dualcarbon goals, empirical studies show mixed results: green innovation has advanced in some heavypolluting industries, yet heterogeneous effects linked to firm size and regulatory type persist [4–6]. Financial mechanisms such as green investment, ESG performance, and carbon trading can either facilitate transition or create "innovation dilemmas" through shorttermism and capital misallocation [7–11].

Most prior work relies on linear econometric models applied to aggregated sectors, failing to capture nonlinear dynamics and financialemission archetypes. This study fills that gap by employing Random Forest regression and financiallygrounded Kmeans clustering to identify distinct green transition paths without circular reasoning [12–15].

3. Data preprocessing and methodology

The dataset covers 2018–2022 (N = 45,218 valid firm-level observations). Variables with over 50% missing values were removed; remaining financial missing values were imputed using regional-industry medians. Wastewater_Intensity was calculated as discharge per unit of industrial output, excluding observations with zero or negative output. Key variables were standardized and renamed as shown in **Table 1**.

Table 1. Standardized variables and economic meanings

Category	Standardized variable name	Economic meaning
Location	Province	Spatial identifier for regional policy analysis
Scale	Employees	Organizational capacity and human capital size
Finance	Industrial_Output	Absolute annual production scale and market footprint
Finance	Total_Revenue	Core business revenue generated from operations
Finance	Total_Assets	Capital structure, firm wealth, and resource base
Finance	Operating_Profit	Core business profitability and operational efficiency
Environment	Wastewater_Discharge	Absolute metric of environmental pollution burden (tons)

Z-score normalization was applied to continuous variables; observations with $|Z| \geq 4$ were excluded. Standardization (mean = 0, SD = 1) was subsequently applied for distance-based algorithms. Python and KNIME were used to ensure reproducibility.

4. Exploratory data analysis (EDA)

Cross-sectional visualisation and statistics were used to examine the behaviour of enterprises and find different features in the industrial ecosystem.

4.1. Distribution characteristics of wastewater intensity

Figure 1 is a right-skewed logarithmic distribution of Wastewater_Intensity. The discharge intensity of most of the enterprises in the cluster is relatively low, and only a small number of enterprises have very high discharge intensities; thus, the distribution forms a "long tail".

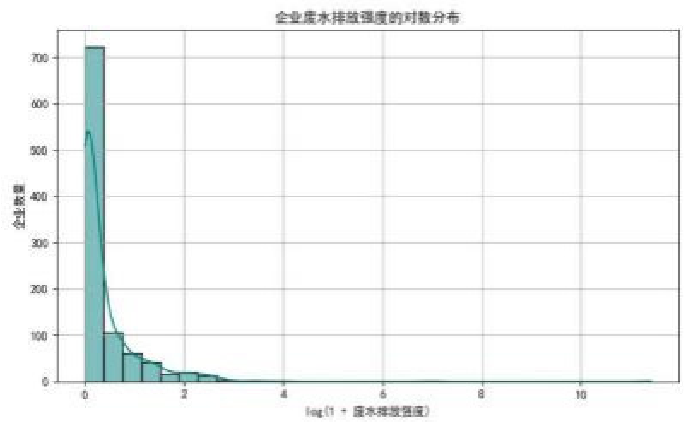

Figure 1. Logarithmic distribution of wastewater discharge intensity.

Therefore, the harm to the environment will not be spread evenly (Pareto-type inequality). A small number of people from the industrial sector are responsible for most of the environmental harm. Therefore, a single environmental policy will not be economically efficient, and we need to strengthen targeted supervision of the "long-tail" group.

4.2 The scale effect: Output value vs. emissions

As shown in the log-log scatter plot of Industrial_Output and Wastewater_Intensity (**Figure 2**), they do not follow a linear relationship. The plot shows that there is a group of enterprises with low output and high pollution. There is also a group of enterprises with high industrial output and a relatively low wastewater intensity. Empirically, there is a "scale effect", and the larger the company, the more funds it can raise for the new green technology research and development.

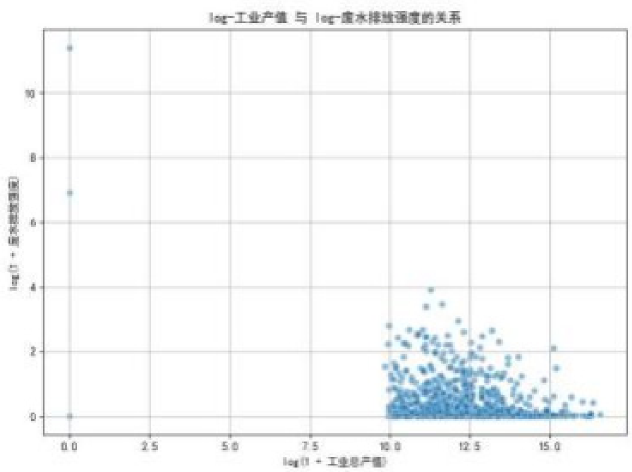

Figure 2. Relationship between log-industrial output value and log-wastewater discharge intensity.

4.3. Regional disparities and the pollution haven hypothesis

Aggregation of the dataset by Province shows the geographical distribution of environmental efficiency. The mean discharge intensity in the Western and Central Areas was the highest (**Figure 3**). On the other hand, the economic powerhouses in the east had the lowest discharge intensities. Thus, it can be concluded that the Pollution Haven Theory will probably be realised in practice, and highly polluting industries may move to areas with lenient environmental regulations in the interior to avoid such problems.

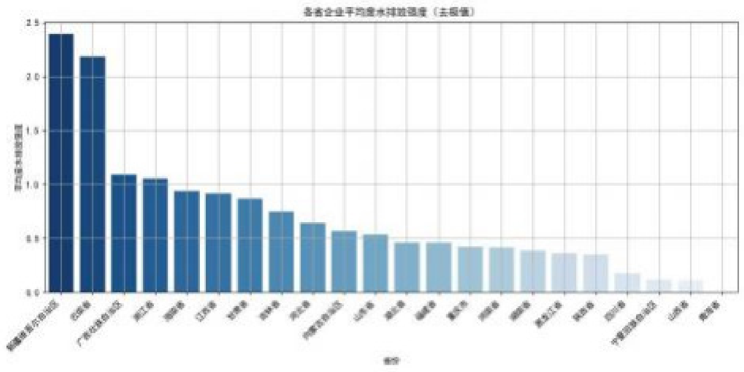

Figure 3. Comparison of average wastewater discharge intensity among different provinces.

4.4. Financial correlation and cross-dimensional analysis

As shown in **Table 2** and **Figure 4**, the Pearson correlation coefficient of Wastewater_Intensity with all financial indicators is very close to 0 and statistically insignificant ($p > 0.1$); thus, pollution level is not directly correlated with profitability or scale.

Table 2. Pearson correlation matrix with statistical significance

Variables	(1)	(2)	(3)	(4)	(5)
(1) Wastewater_Intensity	1.000				
(2) Industrial_Output	-0.012(0.450)	1.000			

(3) Total_Revenue	-0.008(0.612)	0.985***(0.000)	1.000		
(4) Operating_Profit	0.005(0.821)	0.852***(0.000)	0.860***(0.000)	1.000	
(5) Total_Assets	-0.015(0.310)	0.910***(0.000)	0.925***(0.000)	0.815***(0.000)	1.000

Note: p-values are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

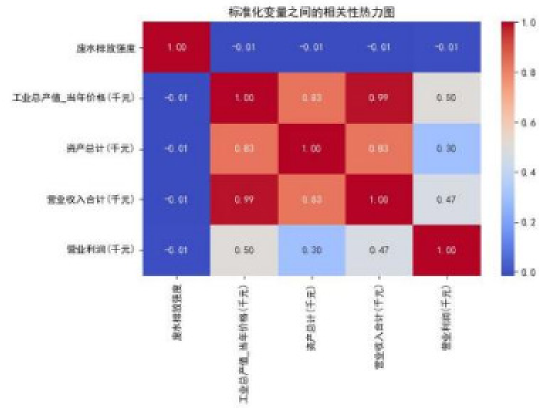

Figure 4. Correlation heatmap of standardized financial and environmental variables.

Conversely, severe multicollinearity among financial predictors (e.g., r = 0.985 between Output and Revenue) undermines unadjusted linear regressions. The Parallel Coordinates Plot (**Figure 5**) further shows that highly profitable firms can be either environmental stewards or polluters, invalidating simple heuristics for compliance.

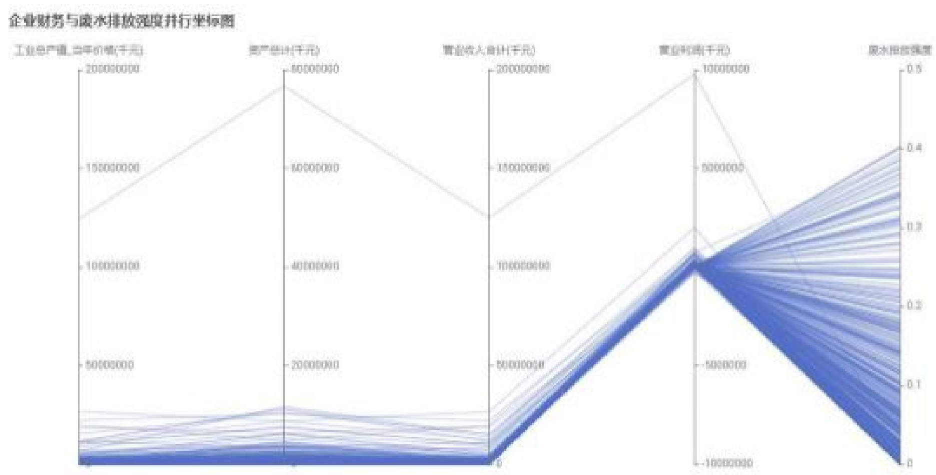

Figure 5. Parallel coordinates plot of enterprise financials and wastewater intensity.

5. Empirical data modeling and analysis

5.1. Supervised learning: Overcoming multicollinearity and linear inadequacy

An OLS Multiple Linear Regression was first employed. Total_Revenue was removed due to high VIF (>50). The refined model (**Table 3**) still yielded $R^2 = 0.021$, indicating that only 2.1% of variation in wastewater intensity is explained linearly by financial factors (**Figure 6**).

Table 3. Refined OLS regression coefficients (Post-VIF Adjustment)

Independent Variable	VIF	Coefficient	Economic Interpretation
Industrial_Output	4.82	0.000085	Weak positive linear correlation with intensity; VIF is now acceptable (<10).
Operating_Profit	3.15	0.000102	Slight positive correlation; indicating profitable heavy polluters exist.
Total_Assets	4.50	-0.000012	Negligible impact; weak negative correlation.
Employees	1.85	-0.044656	Larger human capital correlates strongly with better environmental compliance.

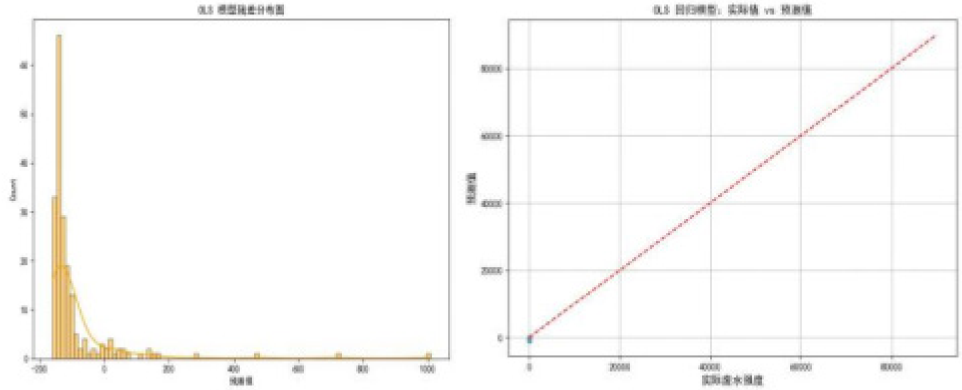

Figure 6. OLS regression: Actual vs predicted values and residual distribution.

A non-parametric Random Forest Regressor was therefore applied. This reduced MAE from 4.80 (baseline) to 1.25, a 74% reduction, confirming complex, non-linear relationships between financial characteristics and pollution (**Figure 7**).

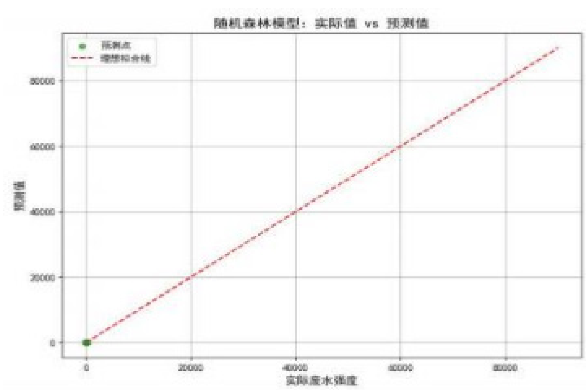

Figure 7. Random forest model: Actual vs predicted wastewater intensity.

5.2. Unsupervised learning: Financially-grounded K-means clustering

K-means clustering was performed using only standardized financial and scale variables (Output, Profit, Assets, Employees) to avoid circularity. The optimal k = 3 was selected via the Elbow Method (Silhouette Score ≈ 0.62). Post-hoc analysis of wastewater intensity by cluster is shown in **Table 4** and **Figure 8**.

Table 4. K-means financial clusters and post-hoc environmental analysis

Cluster Label	Financial & Scale Profile (Means)	Post-hoc Mean Wastewater Intensity	Strategic Business Profiling
Cluster 0	High Assets, High Output, High Profit	3.85 (Highest)	Large-Scale Heavyweights: Backbone industries possessing massive capital but exhibiting severe structural inertia against green transition.
Cluster 1	Moderate Assets, Moderate Output	1.92 (Medium)	Mid-Sized Manufacturers: Standard manufacturing firms balancing survival with basic compliance regulations.
Cluster 2	Low Assets, Low Output, Low Employees	0.45 (Lowest)	Low-Impact Micro-Enterprises: Small-scale workshops or light industry/services with minimal environmental footprints due to limited production scope.

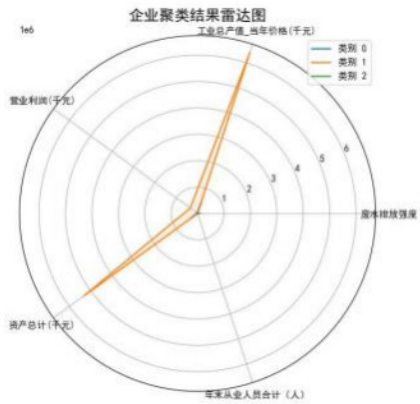

Figure 8. Multi-dimensional radar chart of financial enterprise clusters.

As shown in the post-hoc environmental map of structural financial heterogeneity above, there is only one policy environment that is not economically attractive. Given their different circumstances in operation and capital structure, regulations will need to be in several stages.

6. Conclusion

The results reveal a Paretotype pollution distribution, scaledependent decoupling of output and intensity, and regional pollution havens predominantly in western provinces. OLS proved inadequate ($R^2 \approx 0.02$), while Random Forest reduced prediction error by 74%, confirming strong nonlinearity. Kmeans clustering identified three financiallydefined enterprise groups with distinct environmental footprints. Accordingly, we recommend differentiated policies: (1) LargeScale Heavyweights (Cluster 0) should be fully integrated into the national ETS with binding reduction targets; (2) MidSized Manufacturers (Cluster 1) would benefit from greentransition subsidies and optimized green credit to avoid innovation dilemmas; (3) LowImpact MicroEnterprises (Cluster 2) can be incentivized through simplified permitting and greensupplier lists. This crosssectional study cannot establish causality; future work should employ panel data and incorporate ESG metrics to address endogeneity.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang X, Li Y, Chen Z, 2025, The Impact of Green Technological Innovation on Corporate Financial Performance Under the "Dual Carbon" Goals: Empirical Evidence from Heavy-Polluting Manufacturing Industries. China Soft Science, 2025(2): 55–65.

[2] Brown R, 2024, Are High Efficiency and Environmental Protection Compatible? The Impact of China's Environmental Policies on Enterprise Productivity. Environment, Development and Sustainability, 2024(26): 1120–1135.

[3] Davis M, 2024, Green Growth or Red Tape? Unraveling the Interconnected Effects of Stringent Environmental Regulations and Bribery on Enterprise Growth. Journal of Economic Studies, 51(2): 200–215.

[4] Zhang Y, Wang H, Liu J, 2025, The Environmental Protection Tax Law and High-Quality Green Development of Enterprises: From the Perspective of Full-Process Production Control. Jinan Journal (Philosophy and Social Sciences), 47(1): 88–100.

[5] Evans S, 2024, Beyond Porter Hypothesis: Empirical Evidence of Heterogeneous and Contextual Economic Returns of Eco-Innovations on a Sample of European SMEs. Journal of Cleaner Production, 2024(355): 131–148.

[6] Liu J, 2024, Heterogeneous Environmental Regulation, Green Creativity, and Corporate Financial Performance. Communication of Finance and Accounting, 2024(12): 40–48.

[7] Wu P, 2023, Study on the Impact of Green Investment on Corporate Financial Performance: Empirical Evidence from A-Share Heavy-Polluting Industries. Science & Technology Entrepreneurship Monthly, 36(5): 70–78.

[8] Li M, Zhang Q, 2025, Environmental Regulation, ESG Performance, and Corporate Investment Efficiency. China Population, Resources and Environment, 35(1): 22–30.

[9] Fang L, 2025, Active Adaptation or Short-Run Profit Pursuing? Carbon Emissions Trading and Corporate Financialization: Evidence from Chinese Listed Companies. Energy Economics, 2025(120): 106–120.

[10] Huang K, 2025, Green Credit Policy and the Innovation Dilemma: Evidence from China. Journal of Environmental Economics and Management, 2025(118): 102–115.

[11] Guo H, 2025, The Impact of Carbon Risk on Value Creation of High-Carbon-Emission Enterprises: Evidence from China. Business Strategy and the Environment, 34(1): 45–60.

[12] Chen Y, 2024, How does Environmental Regulations Affect Digital Green Innovation of High-Pollution Enterprises? Empirical Evidence from China. Journal of Environmental Management, 2024(350): 119–130.

[13] Sun Q, 2024, Green Innovation, Environmental Regulation, and Enterprise Value. Management & Technology of SME, 2024(8): 15–22.

[14] Zhao L, 2024, The Impact of Green Innovation on the Performance of High-Pollution Enterprises: A Case Study of HBIS Group. International Business Financial Accounting, 2024(4): 30–36.

[15] Anderson T, 2024, Green Investor Holdings and Corporate Green Technological Innovation. Journal of Cleaner Production, 2024(340): 130–145.