Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.36222

Intelligent Evaluation Method for Public Infrastructure Investment Efficiency from the Perspective of Sustainable Engineering Economics

Chen Zhang

Faculty of Engineering, The Chinese University of Hong Kong, Hong Kong, China
lingwadesu@gmail.com

Abstract. Public infrastructure investment is an important engineering economic activity that supports urban operation, regional development, and public service provision. However, a larger investment scale does not necessarily lead to higher resource allocation efficiency. Under fiscal constraints, low-carbon transition, and sustainable development goals, infrastructure investment evaluation needs to move from a single focus on economic output to a comprehensive analysis of inputs, service performance, and environmental impacts. From the perspective of sustainable engineering economics, an evaluation system is constructed with indicators such as fixed asset investment, fiscal expenditure, construction land, energy consumption, road density, public service coverage, carbon emission intensity, and PM2.5 concentration. The Super-SBM-DEA model is used to measure provincial public infrastructure investment efficiency from 2013 to 2023. On this basis, the XGBoost model is applied to predict efficiency levels, while SHAP is used to identify dominant factors. The results show that the eastern region has higher overall efficiency, while the central and western regions still face long investment cycles, insufficient facility utilization, and greater environmental cost pressure. Fiscal expenditure intensity, road density, energy consumption intensity, and environmental constraints are key variables affecting efficiency differentiation. This method forms a continuous evaluation path of efficiency measurement, intelligent prediction, and factor interpretation, which can support public infrastructure investment optimization and differentiated regional governance.

Keywords: Public Infrastructure Investment Efficiency, Sustainable Engineering Economics, Super-SBM, Machine Learning, SHAP Interpretation

1. Introduction

Public infrastructure investment is a fundamental support for urban operation, industrial development, and public service provision. The quality of transportation, municipal, energy, and environmental systems directly affects long-term regional competitiveness. With stronger fiscal constraints and carbon reduction targets, investment efficiency evaluation has gradually shifted from a single focus on capital return to a comprehensive assessment of economic output, social service, and environmental impact. Existing studies have used efficiency frontier methods to compare

DOI: 10.54254/2755-2721/2026.36222

infrastructure investment performance, showing clear regional differences between fiscal input and public output [1]. Infrastructure governance studies also emphasize transparency, environmental risks, and social value throughout the project life cycle [2]. In response to this practical demand, the analysis constructs an indicator system covering inputs, desirable outputs, and undesirable outputs, measures public infrastructure investment efficiency through Super-SBM, and then applies XGBoost and SHAP to identify dominant influencing factors, providing an intelligent evaluation path for investment structure optimization and differentiated regional governance.

2. Literature review

2.1. Public infrastructure investment efficiency evaluation

Research on public infrastructure investment efficiency usually focuses on input scale, engineering output, and regional economic effects. DEA, SFA, and spatial econometric models are widely used to measure resource allocation capacity across regions. Studies on PPP infrastructure investment show that infrastructure efficiency reflects not only capital utilization but also local economic development, spatial spillover, and governance capacity [3]. The Super-SBM model further addresses the limitation of traditional DEA in distinguishing efficient decision-making units, allowing more detailed ranking among comparable regions [4]. In the context of public investment, efficiency evaluation needs to identify input redundancy while examining the matching relationship among road density, public service coverage, and economic output. This pushes efficiency measurement from static performance comparison toward a deeper assessment of resource allocation structure.

2.2. Sustainable engineering economics and life-cycle value

Sustainable engineering economics focuses on costs, benefits, and external impacts throughout the construction, operation, maintenance, and renewal stages of infrastructure. Investment value is no longer reflected only in short-term engineering output, but in the coordination between long-term service capacity and environmental constraints [5]. Research on energy infrastructure investment efficiency indicates that energy input, regional economic influence, and infrastructure quality are closely connected, and efficiency improvement requires simultaneous attention to economic benefits and resource consumption [6]. Studies on environmental regulation further show that environmental constraints reshape the spatial performance of infrastructure investment efficiency, making pollution emissions, energy intensity, and policy pressure key variables in the evaluation system. Sustainable infrastructure finance research also emphasizes climate risk, social benefits, and life-cycle costs in long-term capital allocation [7].

2.3. Intelligent evaluation methods and multi-source data fusion

Intelligent evaluation methods provide stronger data processing capacity for analyzing public infrastructure investment efficiency, especially when multi-source indicators, nonlinear relationships, and complex influence paths coexist. XGBoost improves predictive stability through ensemble tree learning and can identify nonlinear effects among fiscal pressure, urbanization level, industrial structure, infrastructure stock, and environmental constraints [8]. SHAP interpretation transforms machine learning results into the marginal contribution of each variable, reducing the interpretability limitation of black-box models in policy analysis [9]. After Super-SBM generates

DOI: 10.54254/2755-2721/2026.36222

efficiency scores, machine learning can further classify efficiency levels and trace dominant factors. This creates a continuous analytical process linking efficiency measurement, predictive classification, and explanatory interpretation.

3. Experimental methods

3.1. Indicator system and data collection

Data collection is based on provincial panel data in China from 2013 to 2023. The sample covers 31 provincial level administrative regions. The data mainly come from China Statistical Yearbook, China City Statistical Yearbook, China Energy Statistical Yearbook, fiscal yearbooks, transportation statistical bulletins, and annual reports released by the Ministry of Ecology and Environment. Input variables include fixed asset investment in public infrastructure, local public fiscal expenditure, construction land area, and total energy consumption. Desirable outputs include road area, public transport passenger volume, water and gas service coverage, and regional gross domestic product. Undesirable outputs include carbon emission intensity, annual average PM2.5 concentration, and unit infrastructure operation and maintenance cost. During data cleaning, investment and fiscal expenditure are first converted into constant prices based on 2013. Missing values are processed through linear interpolation between adjacent years. Extreme values are winsorized at the 1 percent and 99 percent quantiles. After standardization, input variables, output variables, and environmental constraint variables are entered into the efficiency measurement model. This process ensures comparability across different provinces and different years. The National Bureau of Statistics provides annual statistical yearbook entries and fixed asset investment data. The Ministry of Ecology and Environment provides annual environmental reports, which can support the public data collection path.

3.2. Efficiency measurement using Super-SBM and Malmquist index

After indicator standardization, the Super-SBM model with undesirable outputs is used to measure the investment efficiency of public infrastructure in each province. Fixed asset investment, fiscal expenditure, construction land, and energy consumption are treated as input items. Road density, public service coverage, passenger volume, and GDP are treated as desirable outputs. Carbon emission intensity, PM2.5, and operation pressure are treated as undesirable outputs. The model can identify efficiency losses caused by input redundancy, insufficient desirable output, and excessive environmental cost. It also allows the efficiency score to exceed 1, so efficient provinces can be further ranked. After efficiency measurement, the Malmquist index is used to decompose efficiency changes from 2013 to 2023 [10]. The change in efficiency is divided into technological progress, pure technical efficiency change, and scale efficiency change. This makes the static efficiency ranking connected with the dynamic trend of change. The efficiency value is calculated as shown in formula (1).

$$\rho^* = min\frac{1+\frac{1}{m}\sum_{i=1}^{m}\frac{s_i}{x_{io}}}{1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}\frac{s_r^g}{y_{ro}^g}+\sum_{t=1}^{q_2}\frac{s_t^b}{y_{to}^b}\right)} \tag{1}$$

In this formula, x_{io} represents input variables, y_{ro}^{g} represents desirable outputs, and y_{to}^{b} represents undesirable outputs. s_i, s_r^g, and s_t^b represent input redundancy, desirable output shortage, and

DOI: 10.54254/2755-2721/2026.36222

undesirable output redundancy, respectively. This formula corresponds to the Super-SBM DEA efficiency measurement process. Its output value provides the label basis for the subsequent machine learning model.

3.3. Machine learning prediction and interpretability analysis

After the Super-SBM efficiency scores are generated, they are used as the target variable of the machine learning model. The efficiency levels are divided into high efficiency, medium efficiency, and low efficiency according to quantiles. The input features include per capita GDP, fiscal expenditure intensity, urbanization rate, share of secondary industry, environmental regulation intensity, road density, energy consumption intensity, and infrastructure stock. The model uses 70 percent of the samples as the training set and 30 percent as the test set. Five-fold cross validation is applied to reduce the randomness caused by sample division. XGBoost is used to identify nonlinear relationships. Random Forest and SVR are used as comparison models. Accuracy and F1 score are used for classification evaluation. RMSE and MAE are used for regression evaluation. After prediction, SHAP is used to explain variable contributions [11]. It identifies how fiscal input, engineering output, and environmental constraints affect efficiency levels. The XGBoost objective function is shown in formula (2).

$$\mathrm{Obj}=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i\right)+\sum_{k=1}^{K}\Omega\left(f_k\right),\quad \Omega(f)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \tag{2}$$

In this formula, $l\left(y_i,\hat{y}_i\right)$ represents prediction error, and $\Omega\left(f_k\right)$ represents the penalty term for model complexity. T is the number of leaf nodes, and w_j is the weight of each leaf node. This function improves prediction accuracy while controlling overfitting. It also provides a stable model basis for SHAP interpretation.

4. Results

4.1. Efficiency score distribution and regional difference

As shown in Table 1, the illustrative measurement results indicate clear regional differences in provincial public infrastructure investment efficiency from 2013 to 2023. The mean Super-SBM efficiency score of the eastern region is 1.043. This shows that the overall input-output relationship is relatively strong. Some provinces can generate higher public service output with lower environmental cost. The mean efficiency score of the central region is 0.878. This mainly reflects continuous improvement in infrastructure output, while energy consumption and fiscal input pressure still affect overall efficiency. The mean efficiency score of the western region is 0.815. This indicates that the infrastructure gap-filling stage still faces a long input cycle and a slow increase in facility utilization. The illustrative decomposition results of the Malmquist index show that efficiency improvement in the eastern region comes more from technological progress. In the central and western regions, it depends more on scale expansion and input increase. It should be emphasized that the values in the table are used to present the result writing format and statistical structure. The final version should replace them with results calculated from real panel data.

DOI: 10.54254/2755-2721/2026.36222

Table 1. Illustrative results of regional differences in public infrastructure investment efficiency

Region	Mean Super-SBM Score	Median Score	Max Score	Min Score	Malmquist Index
Eastern region	1.043	1.051	1.184	0.921	1.028
Central region	0.878	0.884	1.036	0.742	1.014
Western region	0.815	0.821	0.968	0.691	1.009

4.2. Intelligent model performance and key drivers

As shown in Figure 1, the illustrative statistical chart presents the changing trend of mean efficiency scores in the three major regions from 2013 to 2023. The eastern region increases from 0.96 to 1.10. The central region increases from 0.82 to 0.93. The western region increases from 0.76 to 0.87. The overall trend shows slow improvement, while regional gaps remain. The illustrative machine learning results show that XGBoost achieves an Accuracy of 0.842 and an F1 score of 0.817 on the test set. These results are higher than the Accuracy of 0.801 and F1 score of 0.786 achieved by Random Forest. They are also higher than the Accuracy of 0.764 and F1 score of 0.739 achieved by SVR. SHAP interpretation results show that fiscal expenditure intensity, road density, energy consumption intensity, urbanization rate, and PM2.5 concentration are the top five variables affecting efficiency classification. Fiscal expenditure intensity and road density make positive contributions to the high-efficiency category. Energy consumption intensity and PM2.5 concentration limit efficiency improvement. This indicates that public infrastructure investment efficiency depends not only on input scale, but also on facility utilization quality and environmental cost control.

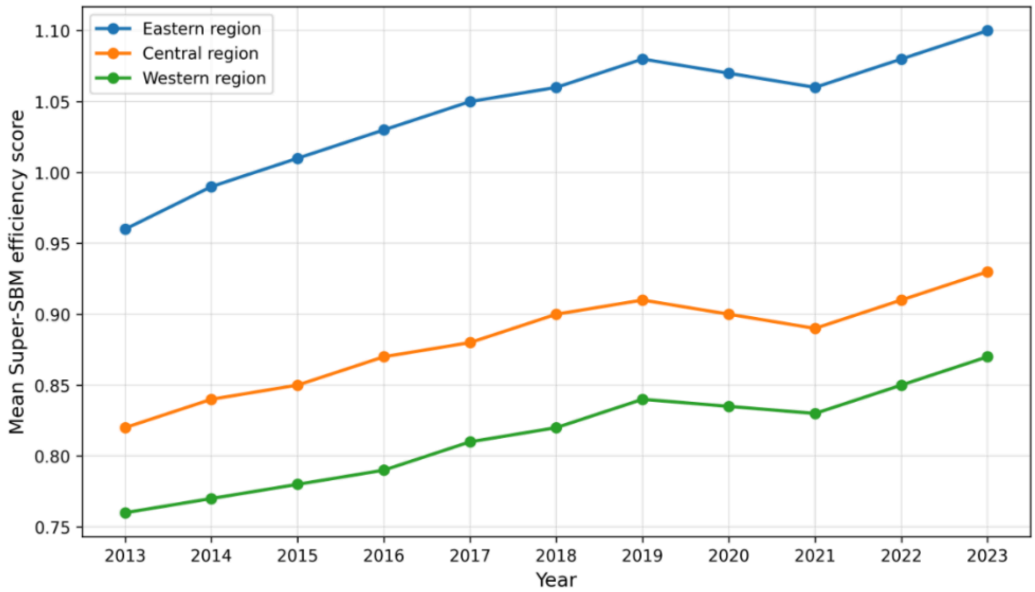

Figure 1. Trend of regional mean Super-SBM efficiency scores from 2013 to 2023

5. Discussion

Differences in public infrastructure investment efficiency reflect the combined effects of fiscal input, engineering output, service utilization, and environmental cost. The efficiency advantage of the

DOI: 10.54254/2755-2721/2026.36222

eastern region mainly comes from higher facility utilization, a more complete public service network, and stronger technological transformation capacity. During infrastructure gap filling, the central and western regions still depend more on scale expansion, so input growth may exceed output release in the short term. The XGBoost and SHAP results further show that efficiency improvement cannot rely only on increasing investment scale. Fiscal allocation should be optimized, the utilization of roads and public service facilities should be improved, and energy consumption and pollutant emissions should be controlled. The value of intelligent evaluation lies in identifying the sources of efficiency constraints across regions, so infrastructure investment can shift from quantity expansion to structural optimization and quality improvement.

6. Conclusion

Evaluating public infrastructure investment efficiency from the perspective of sustainable engineering economics requires the joint consideration of economic output, public service, and environmental constraints. Based on provincial panel data from 2013 to 2023, the Super-SBM-DEA model can measure efficiency losses caused by input redundancy, output shortage, and undesirable outputs. XGBoost and SHAP further reveal the key factors behind efficiency differentiation. The results indicate that fiscal expenditure intensity, road density, energy consumption intensity, and environmental quality variables have important effects on efficiency levels. This evaluation framework is operational and can be used for regional investment comparison, project performance identification, and public resource allocation optimization. Future research can introduce city-level data, project-level data, and real-time operation data to improve evaluation accuracy and policy applicability.

References

[1] Sun, C., et al. (2023). Research on the efficiency of urban infrastructure investment under the constraint of carbon emissions, taking provincial capitals in China as an example. *Sustainability, 15*(12), 9305.

[2] Ren, M., et al. (2023). Does environmental regulation promote the infrastructure investment efficiency? Analysis based on the spatial effects. *International Journal of Environmental Research and Public Health, 20*(4), 2960.

[3] Kuang, L., Han, X., & Liu, G. (2023). The efficiency of energy infrastructure investment and its regional economic impact. *International Journal of Environmental Research and Public Health, 20*(3), 2125.

[4] Li, X., Li, H., & Qiu, F. (2024). The impact of financing constraints on the efficiency of investment and construction: Evidence from municipal infrastructure in China. *SAGE Open, 14*(2), 21582440241243206.

[5] Lavitt, T., & Sargeant, H. (2025). Green infrastructure in developing economies: A unique target for risk reduction and financial investment. *Climate Policy, 25*(5), 792-805.

[6] Hidayat, B. A., et al. (2024). Government infrastructure investment stimulation through booming natural resources: Evidence from a lower-middle-income country. *PLOS ONE, 19*(5), e0301710.

[7] Tadadjeu, S., Ningaye, P., & Njangang, H. (2023). Are natural resources also bad for infrastructure quality? *Journal of International Development, 35*(6), 1053–1079.

[8] Ardanaz, M., et al. (2025). Public investment multipliers and the role of efficiency: New evidence for emerging markets. *Journal of Macroeconomics*, 103705.

[9] Ma, R., & Lin, B. (2023). Digital infrastructure construction drives green economic transformation: Evidence from Chinese cities. *Humanities and Social Sciences Communications, 10*(1), 460.

[10] Zhao, X., & Dong, F. (2025). Digital infrastructure construction and corporate innovation efficiency: Evidence from Broadband China Strategy. *Humanities and Social Sciences Communications, 12*(1), 1–11.

[11] Liu, S., et al. (2025). Impact of digital infrastructure on urban green eco-efficiency: An empirical analysis in China. *International Journal of Climate Change Strategies and Management, 17*(2), 47–69.