Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/92/20241689

Using super-efficiency slack-based measure data envelopment analysis with enhanced “neural network frontier” to evaluate the operating efficiency of Chinese-listed real estate companies

Wenwei Chen

Department of Economics, Jinan University, Guangzhou, 511400, China

a252868@stu2021.jnu.edu.cn

Abstract. China's real estate industry has faced significant changes and challenges in the aftermath of the COVID-19 pandemic and recent years, and has been unable to maintain its former prosperity. Facing a rapidly changing market environment and new loan restriction policies, real estate companies are encountering unprecedented challenges. Efficiency evaluation and analysis of real estate companies are essential to guide them in improving operational efficiency and adapting to new development models. This paper examines 87 real estate development companies listed on the Shenzhen and Shanghai markets, first evaluating and analyzing their operational efficiency in 2023 using the output-oriented super-efficiency slack-based measure data envelopment analysis (SBM-DEA) method. To address the issue of overestimated operational efficiency in the original model, this paper integrates it with the backpropagation neural network (BPNN) algorithm to obtain more robust re-evaluated efficiency scores. The results align with expectations, showing an overall decline in efficiency scores, and more significant decreases for companies with smaller output shortfalls in the original evaluation. The re-evaluated efficiency scores do not excessively underestimate efficiency and can serve as a reliable proxy for the original scores. The enhanced model combining BPNN with super-efficiency SBM DEA provides reliable and robust efficiency evaluations for these 87 listed real estate companies.

Keywords: Data envelopment analysis, Super-efficiency SBM, Neural network, Real estate.

1. Introduction

The real estate industry is crucial to China's economy, serving as a pillar industry and playing a pioneering role during rapid urbanization. After the Reform and Opening-up, China transitioned from a rural society to an urbanized one, with housing demand outpacing supply. From 1998 to 2020, the average price of commercial housing nationwide increased from 20k yuan to 100k yuan. Before COVID-19, the real estate market significantly boosted economic growth, with development investment contributing over 10% to economic growth for many years [1].

The COVID-19 epidemic marked a turning point, causing a shock to real estate demand and construction. The suspension of property sales increased capital chain tensions, leaving projects abandoned. Real estate share prices and property values were negatively affected, and the growth rate of national real estate development investment declined significantly [2]. Figure 1 shows the trend, dominated by the continuous fall, of the national real estate climate index from early 2021 to the end of 2023 [3][4]. In August 2020, the "Three Red Lines" financial regulations were introduced to govern corporate debt financing, impacting the high-leverage, high-growth model. Despite evidence of the immediate negative impacts of the policies, some researchers believe these policies will benefit long-term development [5][6][7].

Given these challenges, investigating real estate productivity is essential to provide deeper insights and promote high-quality development. Data Envelopment Analysis (DEA) measures the efficiency of real estate companies. Previous studies have applied DEA in various markets, demonstrating its utility in assessing efficiency [8][9][10][11]. However, the DEA frontier is very sensitive to the presence of outliers and statistical noise [12], the efficient frontier derived from DEA may be warped if the data are contaminated by statistical noise. Besides, DEA can hardly be used to predict the performance of other decision-making units [13]. Neural Networks (NN) address these limitations by solving non-linear problems and creating a smooth, continuous frontier [14]. Integrating DEA with NN improves robustness, Patel (2015) [15] used NN to smooth DEA's frontier, providing a more accurate evaluation of DMUs.

Few studies have applied neural networks to analyze the efficiency of China's real estate enterprises, this paper aims to provide a review and insight into the operating efficiency of 87 listed real estate development companies in China in 2023. The following work in Figure 2 is done in the paper to establish an enhanced frontier to evaluate the efficiency of the selected companies in 2023. Initially a super-efficiency SBM is performed to get the original evaluation of companies, and then a BPNN is trained to fit an enhanced frontier which is eventually used to re-evaluate the efficiency. Compared to the original frontier, the enhanced frontier provides a more robust evaluation of efficiency by capturing the slacks omitted by the super-efficiency SBM DEA.

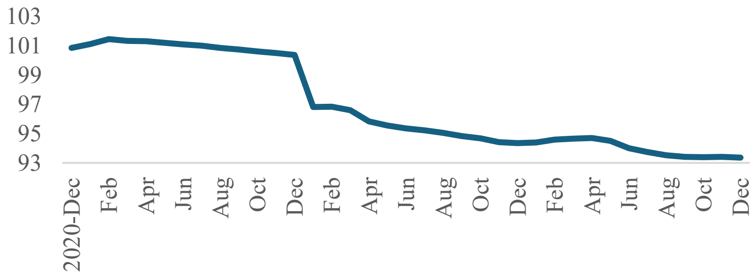

Figure 1. National Real Estate Climate Index (From NBS China)

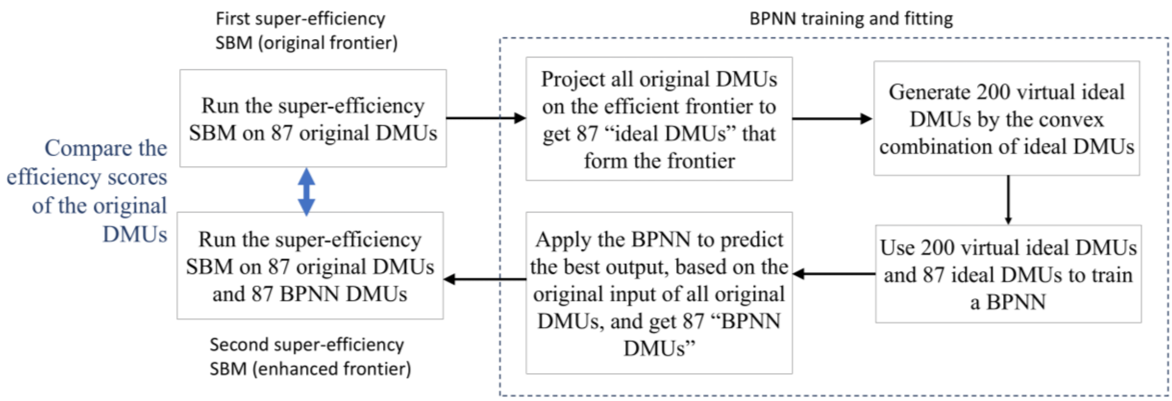

Figure 2. The Construction Process of "Neural Network Frontier"

2. Methodology

2.1. Data envelopment analysis (DEA)

2.1.1. CCR model. The CCR model is a basic DEA model assuming constant returns to scale (CRS) and is radial in nature, as it aims to achieve the greatest proportional reduction in all inputs or expansion in all outputs. The efficiency score of a DMU is derived from how much its inputs can be contracted or outputs expanded proportionately.

The CCR model has two orientations: input- and output-oriented. The orientation corresponds to whether the goal is to reduce excess inputs or expand shortfalls in outputs to move the inefficient unit to the frontier. For instance, in an input reduction model, the greatest percentage reduction in all inputs is sought. The red arrow in Figure 3 illustrates radial improvement for a 2-input, 1-output case. DMU B is diagonally moved to DMU E which lies on the frontier.

To assume there are n DMUs, namely $\mathrm{DMU_j}$ $(j = 1, 2, \dots, n)$, each DMU includes m input variables $X_j = \left(x_{1j}, x_{2j}, \dots, x_{mj}\right)^T$, s output variables $Y_j = \left(y_{1j}, y_{2j}, \dots, y_{qj}\right)^T$, then a output-oriented CCR model of $\mathrm{DMU_{j0}}$ is provided by this linear programming (LP):

- $\max\limits_{\phi,\lambda} \phi$

$$\text{s.t.}\begin{cases} \sum X_j\lambda_j \le X_0 \\ \sum Yj\lambda_j \ge \phi Y_0 \\ \lambda_j \ge 0\ (\forall j) \end{cases} \tag{1}$$

in which (X_0, Y_0) is the input and output of $\mathrm{DMU_{j0}}$, λ_j denotes the intensity of $\mathrm{DMU_j}$ when the linear frontier of efficient DMUs is formed, therefore $\sum X_j\lambda_j$ and $\sum Y_j\lambda_j$ plot the efficient frontier. This formula means that within the production possibility set T, while maintaining output X_0 without increase, the input vector Y_0 is maximized proportionally to ϕY_0. If the output cannot be enlarged, namely the maximum value of ϕ is 1, the evaluated DMU is considered an efficient unit; otherwise, it is relatively inefficient.

2.1.2. Super-efficiency SBM. The CCR model is unable to fully capture the impact of slackness on efficiency. In Figure 3, DMU B is more efficient than DMU E', as DMU B's input2 productivity is higher than that of DMU E' while controlling for input1 productivity. However, both DMUs are identified as efficient by the CCR model. The slack-based measure (SBM) addresses this issue by introducing "slacks", which quantify the inefficiencies by indicating the degree to which inputs can be decreased or outputs can be increased while still maintaining the current level of outputs and inputs, respectively. In SBM, input and output factors do not need to change proportionally, but need to eliminate the slacks. The movement from DMU E' to B which is plotted as a blue arrow in Figure 3 represents a slack improvement, recognizing B as more efficient than E', and E is the least efficient in this 3 DMUs.

When many DMUs are found efficient in a DEA model, the relative efficiency comparison fails as all efficient DMUs score 1. To address this, super-efficiency DEA models can be used. In a super-efficiency model, the DMU being evaluated is excluded from its own comparison set, removing the upper limit of 1 on efficiency scores and allowing scores greater than 1, which can rank efficient units. An output-oriented super-efficiency SBM with constant returns to scale is given by:

- $\max\limits_{\lambda,s^-,s^+} \rho = 1/(1 + \frac{1}{q}\sum_{r=1}^{q}\frac{s_r^+}{y_{io}})$

$$\text{s.t.}\begin{cases} X_0 \geq \sum_{j\neq 0} X_j\lambda_j + s^- \\ Y_0 \leq \sum_{j\neq 0} Y_j\lambda_j - s^+ \\ \lambda_j \geq 0\ (\forall j), s^- \geq 0, s^+ \geq 0 \end{cases} \tag{2}$$

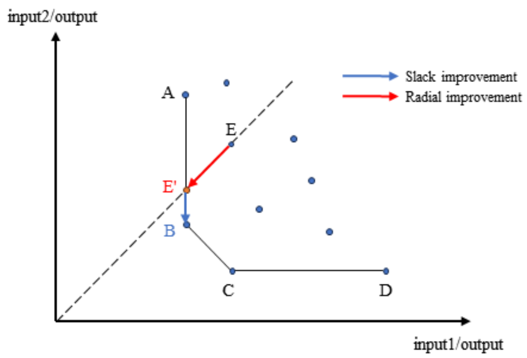

Figure 3. Efficient Frontier and Efficiency Improvement in DEA

2.2. *Neural network*

The neural network is a machine learning algorithm composed of interconnected processing units arranged into layers [16]. Each unit has multiple inputs, each with an associated weight. The unit performs a weighted summation of the inputs, applies a transfer function, and transmits the output to the next layer. A classic model is the backpropagation neural network (BPNN), a multi-layer forward neural network trained using the error backpropagation algorithm. BPNN can solve non-linear problems that simple perceptions cannot. It consists of three layers: input, hidden, and output. Using the gradient descent algorithm, it minimizes the square of the error between the actual and expected outputs.

In this paper, a BPNN with 3 layers and 128 neurons in each layer was trained. The network used 3 input variables as features and 2 output variables as labels, with ReLU for the hidden layers and a linear activation function for the output layer. The architecture of the BPNN is shown in Figure 4.

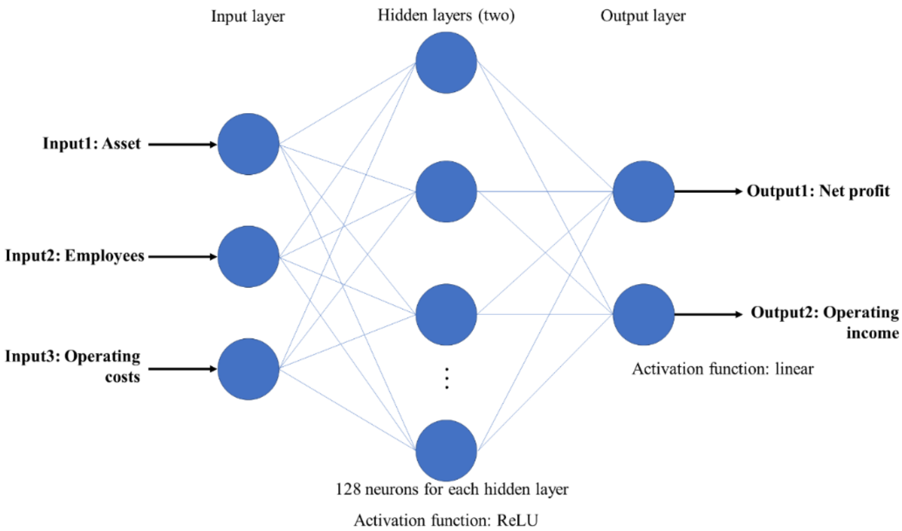

Figure 4. The architecture of BPNN

3. Data selection and descriptive statistics

Data from 87 real estate development companies listed on the Shanghai Stock Exchange or Shenzhen Stock Exchange in 2023 is selected from CSMAR. Table 1 shows the variable selection and denotation.

DOI: 10.54254/2755-2721/92/20241689

Figure 5 plots the distribution of these variables. These variables are relatively concentrated in most companies, with a few larger-scale companies. The number of companies with positive net profit is roughly equal to the number of companies with negative net profit.

Table 1. The Selected Variables for DEA

Category	Variable	Denotation
Input	Total asset	Total of various asset items
	Number of employees	The total number of employees in a listed company, as disclosed in the annual report, refers to the number of registered (on-duty) employees
	Operating cost	Operating costs recognized by the company
Output	Net profit	Operating income recognized during the company's operations
	Operating income	Net profit achieved by the company

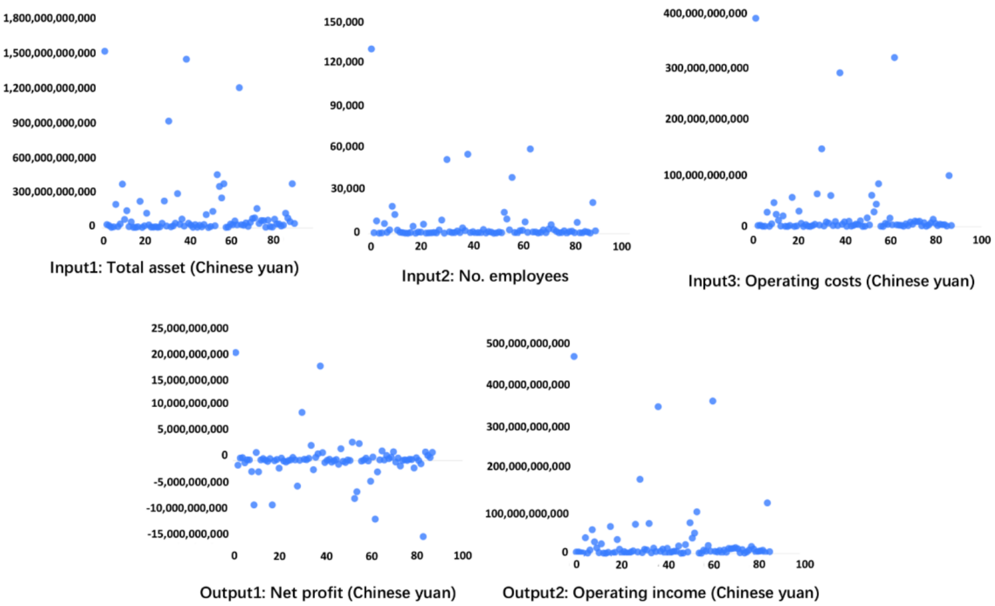

Figure 5. Distribution of the Selected Variables

4. Empirical results analysis

4.1. Original measure of efficiency: Super Efficiency SBM-DEA

The output-oriented super-efficiency SBM DEA is performed on Python to evaluate the operating efficiency of these 87 companies as “original DMUs”. 3 input variables and 2 output variables listed in Table 1are used. To perform the DEA, all variables are Min-Max normalized to scale the variables of this dataset between 0 and 1 for large differences in the original magnitudes which makes the model solution biased or infeasible. A negative net profit is assigned as 0.01. In Tabel 2, the first column shows the 6-digit codes for Shanghai and Shenzhen A-shares assigned to these companies listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. The original efficiency scores of all DMUs are listed in the second column of Table 2 in descending order.

Table 2. The Efficiency Evaluation Results

Code	Original Eff.	Adjusted Eff.	Code	Original Eff.	Adjusted Eff.	Code	Original Eff.	Adjusted Eff.
002244	1.238	0.911	000809	0.973	0.916	000006	0.921	0.875
600639	1.045	0.957	000863	0.973	0.915	000926	0.919	0.849
000517	1.041	1.022	000965	0.972	0.925	000736	0.911	0.859
000014	1.010	0.946	000036	0.972	0.906	600185	0.911	0.861
002133	1.009	0.962	600716	0.971	0.918	600683	0.904	0.865
600663	1.009	0.751	000514	0.971	0.907	000090	0.900	0.743
600007	1.008	0.918	600848	0.970	0.878	600665	0.892	0.849
601155	1.006	0.549	000718	0.969	0.903	600657	0.887	0.816
000797	1.006	0.822	600641	0.969	0.906	600736	0.885	0.836
600208	1.006	0.911	600603	0.967	0.920	002305	0.881	0.833
600743	1.004	0.938	600159	0.966	0.910	600622	0.879	0.833
600173	1.002	0.952	600895	0.966	0.889	600708	0.878	0.840
000897	1.002	0.945	600064	0.964	0.921	002314	0.874	0.810
600724	0.996	0.920	600773	0.962	0.913	000042	0.865	0.807
000838	0.992	0.937	000029	0.959	0.904	000402	0.799	0.713
600094	0.992	0.956	600246	0.959	0.903	000031	0.771	0.711
600692	0.982	0.919	600322	0.957	0.909	600565	0.732	0.658
000573	0.980	0.924	000011	0.956	0.757	600376	0.697	0.629
600234	0.980	0.922	002208	0.953	0.913	600325	0.695	0.619
000668	0.980	0.924	600067	0.949	0.910	000961	0.605	0.574
600082	0.979	0.926	600162	0.945	0.855	000656	0.565	0.522
600503	0.979	0.922	600638	0.945	0.890	600383	0.560	0.479
000886	0.978	0.913	600649	0.941	0.837	600048	0.551	0.500
000631	0.978	0.922	600791	0.940	0.897	600340	0.519	0.443
601512	0.978	0.895	600648	0.937	0.869	001979	0.485	0.425
002016	0.977	0.907	600510	0.927	0.878	000069	0.400	0.337
000609	0.975	0.920	600675	0.925	0.889	000002	0.344	0.330
600748	0.975	0.789	600266	0.925	0.840	600823	0.313	0.274
600463	0.975	0.910	600533	0.923	0.859	600606	0.140	0.128

Descriptive Statistics of Original Efficiency	Descriptive Statistics of Adjusted Efficiency
Max: 1.238	Max: 1.022
Min: 0.140	Min: 0.128
Median: 0.959	Median: 0.89
Mean: 0.890	Mean: 0.817
Standard deviation:0.182	Standard deviation:0.175

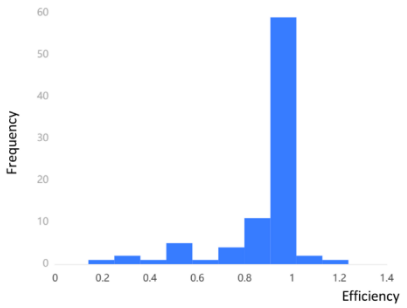

Figure 6. Frequency Distribution of Efficiency in the Original Evaluation

There are 13 efficient DMUs that have an efficiency of not less than 1. 73.5% of these companies have an efficiency higher than 0.9. It was found that most of the DMUs have efficiency scores concentrated between 0.8 and 1. Figure 6 shows that when the efficiency drops from 0.9 to 0.8, the number of DMUs decreases sharply. This evaluation suggests that the majority of the companies exhibit comparable operational efficiency, with little difference from the top-performing companies.

4.2. Revised measure of efficiency: using BPNN to enhance the frontier

Based on super-efficiency SBM DEA, I aim to determine more robust and realistic efficiency scores for DMUs. DEA models often provide exaggerated efficiency scores because efficiency is measured relative to available DMUs. A small proportion of efficient DMUs establishes a rough frontier, potentially overestimating DMU efficiency. Figure 7 illustrates a 2-input, 1-output SBM model with a non-smooth original frontier. X on this frontier is X', and vector XX' can be decomposed into two orthogonal slack vectors. In an output-oriented SBM model, a DMU's efficiency is determined by its slacks; larger output slacks indicate lower efficiency

A non-smooth frontier can underestimate the distance between a DMU and the efficient frontier. If point R does not exist in Figure 7, DMUs A, Q, B, and C form the efficient frontier. When DMU R appears between points A and B, the frontier changes to ARBC. In reality, DMU R represents DMUs missing from the corpus due to a lack of observations and is not an extreme value. DMU R likely belongs to the production possibility set and can form a smoother frontier with points A, B, and C. The slack vectors decomposed from vector PR are larger than the original slack vectors, indicating the original frontier omitted some actual output slack, resulting in underestimated slack and overestimated efficiency for point P.

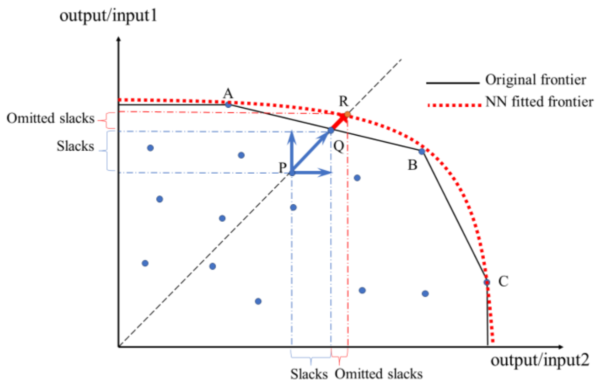

Figure 7. NN enhanced frontier adjusts the underestimated slacks and overestimated efficiency

Based on the above, I used a backpropagation neural network (BPNN) to simulate an enhanced frontier that is smoother by including as many potentially unobserved valid DMUs similar to R as possible in the corpus. This approach provides more robust efficiency estimates for the initial DMUs, addressing the common issue of SBM DEA underestimating slacks.

I firstly projected 87 original DMUs onto the original frontier by deducting slacks from the input variables and adding slacks to the output variables. This process created 87 "ideal DMUs" that accurately represent the original efficiency frontier. These ideal DMUs are all considered efficient. Next, I generated an additional 200 virtual ideal DMUs, each a convex linear combination of all 87 ideal DMUs, ensuring these virtual DMUs are also efficient. Using Python 3.12, I then trained a BPNN with the 87 ideal DMUs and the 200 virtual DMUs. The training used 3 input variables as features and 2 output variables as labels, employing 5-fold cross-validation and training the model for 500 epochs to ensure robustness. This BPNN model effectively learns the relationship between the input and output of efficient DMUs. The BPNN model was then used to predict the optimal outputs for the 87 original DMUs by using their input variables. The predicted best outputs, combined with the original inputs,

DOI: 10.54254/2755-2721/92/20241689

formed a new set of 87 "fitted DMUs." These fitted DMUs simulate the efficient but unobserved DMUs that might exist within the corpus. Finally, all original DMUs and fitted DMUs were applied in Super SBM-DEA. The fitted DMUs formed an enhanced frontier. The efficiency re-evaluation results are shown in the third column of Table 2.

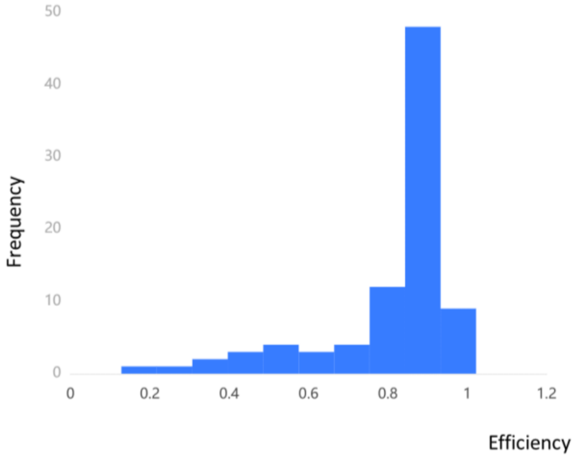

Figure 8. Frequency Distribution of Efficiency in the Adjusted Evaluation

After re-evaluating the efficiency of the original DMUs, it is clear that the re-evaluated efficiency scores for all companies dropped, with an average percentage decrease of 8.34%. The mean efficiency decreased by 8.2% compared to the original evaluation, and 47.13% of the companies have efficiency scores over 0.9. Compared to the initial efficiency distribution in Figure 8, Figure 6 indicates that the efficiency scores in the reevaluation are more dispersed, allowing for better differentiation of efficiency differences between companies. The goal of establishing an enhanced frontier using fitted DMUs to reduce the overestimation of the original DMUs' efficiency has been achieved. The original rough frontier is well enveloped by the enhanced, smoother, and expanded frontier. To further ensure the enhanced frontier's effectiveness, I additionally added the ideal DMUs obtained from the projection in step 1 into another calculation. The efficiency scores of the original DMUs did not change, and the average score of the ideal DMUs that construct the original frontier is 0.940, which is very close to the efficient level. This means the enhanced frontier does not excessively deviate from the original frontier, ensuring that the DMUs on the original frontier and the DMUs enveloped in the original frontier are not excessively underestimated. The re-evaluated efficiency is a robust adjustment to the original efficiency.

To verify the ability of the re-evaluated efficiency to accurately reflect the original efficiency, a regression analysis was conducted. This analysis aimed to determine the extent to which the re-evaluated efficiency can serve as a reliable proxy for the original efficiency scores. The results are shown in Table 3. The regression coefficient is significantly different from zero and close to one, with an R^2 value of 0.995. This indicates that the re-evaluated efficiency can serve as a reliable proxy for the original efficiency.

Table 3. Linear Regression Results of the Reevaluated Efficiency to The Original Efficiency

Original efficiency	Coef.	St. Err.	t-value	p-value	[95% Conf	Interval]	
Reevaluated efficiency	1.09*	0.01	129.12	0	1.07	1.10	
Mean dependent var	0.89		SD dependent var		0.18		
R-squared	1.00		Number of observations		87		
F-test	16672.66		Prob > F		0.00		
*** p<.001, ** p<.01, * p<.05							

To identify which DMUs undergo the largest adjustments in the re-evaluation, I performed a multiple regression analysis of the percentage change in efficiency against the slack variables of the original DMUs from the first DEA calculation. The results are shown in Table 4. It can be observed that DMUs with larger slacks in total assets and number of employees, but smaller slacks in operating income in the initial DEA calculation are more likely to undergo significant adjustments in the reevaluation. This indicates that DMUs with larger input redundancies and smaller output shortfalls are more likely to undergo greater adjustments in the reevaluation. Those DMUs that achieved higher efficiency scores primarily due to underestimated output shortfalls will receive lower scores in the reevaluation. This aligns with the objective of the output-oriented model, which is to guide improvements in output.

Table 4. Linear Regression Results of Slacks to the Percentage of Efficiency Change

Percentage of efficiency change (absolute value)	Coef.	St. Err.	t-value	p-value	[95% Conf	Interval]	
Asset slacks	0.76***	0.22	3.50	0.00	0.33	1.19	
Employee slacks	1.92***	0.39	4.86	0.00	-1.13	-2.70	
Costs slacks	-2.58	1.39	-1.85	0.07	-0.19	5.34	
Profit slacks	0.01	0.02	0.77	0.45	-0.04	0.02	
Income slacks	-4.01**	1.34	-3.00	0.00	1.35	6.67	
Constant	0.08***	0.01	11.87	0.00	-0.09	-0.06	
Mean dependent var	-0.08		SD dependent var		0.06		
R-squared	0.42		Number of observations		87		
F-test	11.74		Prob > F		0.00		
*** p<.001, ** p<.01, * p<.05							

5. Conclusion

This paper examines 87 real estate development companies listed on the Shenzhen and Shanghai markets, analyzing their input and output data from 2023. Firstly, the paper employs the output-oriented super-efficiency SBM DEA method to evaluate the operational efficiency of the selected companies. It finds that the operational efficiency of most companies is above 0.8, with 73.5% of the companies having an operational efficiency greater than 0.9. To address the issue of overestimated DMU efficiency in the original model, this paper integrates the backpropagation neural network (BPNN) algorithm into the output-oriented super-efficiency SBM DEA method. By using this method, the study provides robust efficiency estimates for the selected companies. All companies' efficiency scores decreased compared to the original evaluation, with an average percentage decrease of 8.34%. The mean efficiency in the reevaluation decreased by 8.2% compared to the original evaluation, with 47.13% of the companies having an efficiency greater than 0.9. The overall decrease in efficiency scores after training an enhanced frontier with the BPNN and incorporating it into the efficiency evaluation aligns with the expectations for this "neural network frontier." The neural network frontier method does not excessively underestimate the efficiency of the companies and has good proxy capabilities for the original efficiency scores. The results of efficiency evaluation using the BPNN to establish an enhanced frontier are robust and effective. Additionally, the enhanced frontier significantly corrects the efficiency scores of companies with originally smaller output slacks, effectively identifying and assigning lower efficiency scores to those with underestimated output shortfalls. This combined super-efficiency SBM DEA and neural network efficiency evaluation method has the potential to be further applied to other output-oriented efficiency evaluation problems.

References

[1] Ni, P., Ni, Z., Xu, H. (2023) Research on New Models and Construction of China's Real Estate Development. China Real Estate Finance, 06: 52–59.

[2] National Bureau of Statistics of China. (2022) National Real Estate Development and Sales in 2021. https://www.stats.gov.cn/english/PressRelease/202201/t20220118_1826502.html
[3] National Bureau of Statistics of China. (2023) National Real Estate Development and Sales in 2022. https://www.stats.gov.cn/english/PressRelease/202301/t20230118_1892298.html
[4] National Bureau of Statistics of China. (2024) Investment in Real Estate Development in 2023. https://www.stats.gov.cn/english//PressRelease/202402/t20240201_1947107.html
[5] Chen, Z., Gao, P., Liang, K. (2023) Impact of COVID-19 on China's Real Estate Industry - An Empirical Study Based on the Stock Market. In: 2nd International Conference on Business and Policy Studies. Singapore. pp. 83.
[6] Lin, A., Hale, T., Lockett, H. (2021) Half of China's top developers crossed Beijing's 'red lines.' Financial Times, 9.
[7] Chu, X., Deng, Y., Tsang, D. (2023) Firm leverage and stock price crash risk: The Chinese real estate market and three-red-lines policy. J. Real Estate Finance Econ., 1–39.
[8] Wong, W.P., Gholipour, H.F., Bazrafshan, E. (2012) How efficient are real estate and construction companies in Iran's close economy? Int. J. Strateg. Prop. Manag., 16(4): 392–413.
[9] Chen, Q., Li, F. (2017) Empirical analysis on efficiency of listed real estate companies in China by DEA. iBusiness, 9(3): 49–59.
[10] Zheng, X., Chau, K.W., Hui, E.C. (2011) Efficiency assessment of listed real estate companies: an empirical study of China. Int. J. Strateg. Prop. Manag., 15(2): 91–104.
[11] Atta Mills, E.F.E., Baafi, M.A., Liu, F., Zeng, K. (2021) Dynamic operating efficiency and its determining factors of listed real-estate companies in China: A hierarchical slack-based DEA-OLS approach. Int. J. Finance Econ., 26(3): 3352–3376.
[12] Bauer, P.W. (1990) Recent developments in the econometric estimation of frontiers. J. Econom., 46(1-2): 39–56.
[13] Wu, D.D., Yang, Z., Liang, L. (2006) Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. Expert Syst. Appl., 31(1): 108–115.
[14] Zhong, K., Wang, Y., Pei, J., Tang, S., Han, Z. (2021) Super efficiency SBM-DEA and neural network for performance evaluation. Inf. Process. Manag., 58(6): 102728.
[15] Bose, A., Patel, G.N. (2015) "NeuralDEA"–A framework using Neural Network to re-evaluate DEA benchmarks. OPSearch, 52: 18–41.
[16] Cane, V.R. (1967) Mathematical models for neural networks. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley. pp. 21–26.