

Data Mining Study on Performance of Poverty Alleviation Pairing Assistance Projects Based on DEA Model and Principal Component Analysis

Dongqing Wu^{1*}, Tianbao Liang², Chaolong Zhang¹, Chaojun Hou¹

¹ School of Mathematics and Data Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China

² School of Management, Zhongkai University of Agriculture and Engineering, Guangzhou, China

Email Address

rickwu@zhku.edu.cn (Dongqing Wu)

*Correspondence: rickwu@zhku.edu.cn

Received: 13 September 2023; Accepted: 27 September 2023; Published: 21 October 2023

Abstract:

This paper examines the performance of poverty alleviation pairing projects in City B, supported by City A in Guangdong Province, China, during the period of 2016-2020, using the data envelopment analysis (DEA) model and principal component analysis. It establishes an evaluation index system for the performance of poverty alleviation pairing support projects and conducts data mining research. Through empirical analysis, it identifies deficiencies in the pairing support work and proposes recommendations such as strengthening the presence of village-based officials and increasing investment in industrial poverty alleviation.

Keywords:

Poverty Alleviation Project Performance, Data Mining, DEA, PCA

1. Introduction

Pairing assistance is a regional mutual aid mechanism with distinctive Chinese characteristics and significant advantages [1]. Significant economic disparity exists between the Pearl River Delta (PRD) region and the eastern, western, and northern parts of Guangdong Province, China. To balance the economic gap and achieve common prosperity between the PRD and the eastern, western, and northern regions, Guangdong Province has implemented three rounds of poverty alleviation pairing assistance strategies from 2009 to 2020. Six cities in the PRD region have been paired with 12 cities in the eastern, western, and northern regions of Guangdong Province [2]. Among them, City A in the PRD region has provided over 3 billion yuan in funding to support City B. Pairing assistance projects have played an essential role in promoting regional coordinated development and the development of poverty-stricken areas. Data mining and evaluating the performance of these projects are of great significance for future decision-making.

Project performance evaluation is a common data mining method that refers to the evaluation of the degree of achievement of performance goals determined by certain evaluation methods, quantified indicators, and evaluation criteria, as well as the execution results of budget arrangements to achieve these goals. Many domestic and foreign scholars have studied the construction of project performance evaluation methods and indicator systems. Lin used the analytic hierarchy process (AHP) to determine the weight of each indicator, analyzed the poverty alleviation targeting performance evaluation results after processing the indicator data, and verified the indicator system through empirical research [3]. Zhao constructed a performance indicator evaluation system for poverty alleviation projects in Shandong Province, studied the performance indicator system using the AHP method, and analyzed the input, output, management, and effect of poverty alleviation projects [4]. Zhang used data acquisition, quantitative evaluation, and comparative analysis to mine a multidimensional and compatible rural development model, providing theoretical and methodological support for the performance evaluation of pairing assistance projects [5]. Foreign scholars have also conducted in-depth research on project performance evaluation. For example, literature [6] provides indicators, key performance indicators (KPIs), and dashboard models and methods for project performance evaluation. Literature [7] proposes the concept and method of Earned Value Project Management and explores how to effectively evaluate project performance. Atkinson proposed that the data envelopment analysis (DEA) model can be used to conduct project performance research in project management [8]. In summary, the DEA model, as a non-parametric evaluation method, has been widely used in project performance evaluation research. These research results show that the DEA model can objectively evaluate project performance and provide decision support for project improvement and optimization.

As countries worldwide strengthen their focus on poverty issues, scholars from different fields have made certain achievements in their research on poverty. However, these studies are mostly qualitative and lack quantitative and empirical research. The performance evaluation of pairing assistance projects has always been a significant concern. Overall, the theoretical research on pairing assistance lags behind practical exploration, and there is a need to strengthen basic theoretical and strategic planning research. There are still some issues and shortcomings in domestic and foreign research. Firstly, the performance evaluation index system for pairing assistance projects is not yet perfect and lacks unified standards and guidance. Secondly, existing research mainly focuses on the construction of performance evaluation methods and models, with relatively less evaluation of the actual effects and impacts of pairing assistance projects. Additionally, there is still a gap between theoretical and practical research in the performance evaluation of pairing assistance projects, and further integration of theoretical and empirical research is needed.

Given the aforementioned problems and shortcomings, this paper aims to conduct data mining research on the poverty alleviation pairing assistance projects carried out by City A to City B from 2016 to 2020, based on the DEA model and principal component analysis (PCA) method. The research aims to evaluate the performance of the poverty alleviation pairing assistance projects. The DEA model can assess the efficiency and performance level of the projects, helping to identify existing problems and shortcomings.

2. Evaluation Model and Indicator System

This section explores the significance of evaluation models and indicator systems in project performance evaluation, highlighting their importance in providing insights and guiding decision-making for project improvement and optimization.

2.1.2. Evaluation Model

The effectiveness of poverty alleviation is mainly reflected in per capita disposable income, which corresponds to the output of the project. Various poverty alleviation measures in developed regions supporting impoverished villages, such as policy subsidies, condolences, industrial poverty alleviation, asset poverty alleviation, reward-based incentives, housing renovation, subsidized loans, transfer payments, and photovoltaic poverty alleviation, correspond to inputs. Therefore, according to performance theory [9, 10], studying the efficiency of inputs/outputs of the project can reflect the quality and utility of poverty alleviation results. Based on the analysis of relevant research in the previous sections and in combination with the actual situation in Guangdong, this paper's evaluation model is primarily based on DEA [11] and PCA [12].

2.2. Indicator System

Through the study of poverty alleviation policies at the national and Guangdong provincial levels, we select a set of input and output indicators and constructed an indicator system, as follows:

(1) Input Indicators

These include input funds and input manpower indicators. Input funds reflect the direct assistance efforts of the supporting party and encompass the total sum of funds for various poverty alleviation projects, such as policy subsidies, condolences, industrial poverty alleviation, asset poverty alleviation, reward-based incentives, housing renovation, subsidized loans, transfer payments, and photovoltaic poverty alleviation. Input manpower indicators reflect differentiated precision assessment and include the number of responsible personnel per average impoverished household, the number of business training sessions conducted for poverty alleviation officials and village leaders, and the number of poverty-stricken individuals receiving entrepreneurship and practical skills training.

(2) Output Indicators

These include the annual per capita disposable income indicator and social poverty reduction indicators. The annual per capita disposable income indicator represents the disposable income remaining after deducting personal taxes and social insurance expenses from the annual per capita income. It provides a straightforward reflection of poverty alleviation results and is a critical poverty reduction indicator. Social poverty reduction indicators are based on the social development, transportation, compulsory education, basic medical care, and housing security of impoverished villages. They include the coverage rate of compulsory education, the coverage rate of basic medical care, the coverage rate of safe housing, the increase in village road mileage, and the construction of safe drinking water facilities.

3. Empirical Study

This paper mainly uses three types of data: first, on-site sampling surveys; second, industry sector archives and precision poverty alleviation databases; and third, relevant literature materials. There is some overlap in the data from these three sources, and we used cross-validation to improve the data quality by removing illogical parts. The research tool used was DEAP (Data Envelopment Analysis Program), which is a widely used DEA software package that provides various DEA model calculation and analysis functions [13]. It is developed based on the Python, and has flexibility and scalability.

3.1. DEA Study

For the analysis of poverty alleviation support data from City A to City B during the period of 2016-2020, we will utilize the DEAP software. To begin with, it is necessary to determine the weights of each decision-making unit to achieve the optimal proportion between their inputs and outputs. We will use the CCR (Charnes, Cooper, and Rhodes) model for computation. Firstly, we need to transform the input and output data into a standardized form by dividing each decision-making unit's inputs and outputs by their respective maximum values. Next, the CCR model will be employed for calculation [14]. The CCR model determines the relative efficiency of each unit and assigns a relative efficiency score by solving a linear programming problem. Higher scores indicate higher efficiency levels. The mathematical expression of this model is as follows [15]:

$$\begin{aligned} & \max \theta, \\ & \text{s.t.} \quad \sum_{j=1}^m v_j y_{ij} - \theta \sum_{i=1}^n u_i x_{ij} \geq 0, \quad i = 1, 2, \dots, n, \\ & \quad \sum_{j=1}^m v_j = 1, \\ & \quad u_i \geq 0, \quad i = 1, 2, \dots, n, \\ & \quad v_j \geq 0, \quad j = 1, 2, \dots, m, \end{aligned} \tag{1}$$

where $\theta = \frac{\sum_{j=1}^m v_j y_j^+}{\sum_{i=1}^n u_i x_i^+}$, represents the relative efficiency score of each decision-making unit. It is used to measure the performance level of a unit relative to the optimal efficiency given the inputs and outputs. By maximizing the value of θ , we can find the optimal efficiency score for each unit. Higher values of θ indicate higher levels of efficiency. Integer n denotes the number of decision-making units, m represents the number of input and output indicators, x_{ij} denotes the input of unit i for output variable j , y_{ij} represents the output of unit i for output variable j , x_i^+ and y_j^+ represent the input and output vectors of unit i relative to the optimal efficiency, respectively. u_i and v_j represents the weights assigned to the input and output variables.

We primarily evaluate the performance of the project by calculating the composite efficiency, pure technical efficiency, and scale efficiency. The results obtained are as follows: The effective years of City A's poverty alleviation support to City B are 2016, 2017, 2019, and 2020, during which the composite efficiency, pure technical efficiency, and scale efficiency are all equal to 1. The average composite efficiency in 2018 is 0.971, the average pure technical efficiency is 1, and the average scale

efficiency is 0.902. Overall, the poverty alleviation efficiency is effective. However, the fact that the composite efficiency and scale efficiency in 2018 are not equal to 1 indicates that the main reason for the overall inefficiency of poverty alleviation policies is uneconomical scale, suggesting an unreasonable allocation of resources. The scale returns from 2017 to 2018 are in a decreasing phase, indicating that the poverty alleviation efforts are still insufficient, and increasing investment in poverty alleviation would be beneficial for improving efficiency. The year 2017 shows significant achievements. Although there was a large investment in 2020, the scale returns decreased slightly, indicating increased pressure during the critical stage of poverty alleviation and a tendency of slightly weakened effectiveness.

3.2. PCA Study

PCA is a commonly used technique for data dimensionality reduction and feature extraction. It achieves this by linearly transforming the original data into a new set of variables called principal components, which are linear combinations of the original variables that capture the maximum variance in the data. PCA helps us understand the main patterns of variation in the data and can reduce the dimensionality of the data, simplifying analysis and visualization. In this study, PCA is employed to analyze the output indicators.

(1) Suitability test: Firstly, categorical variables are transformed into continuous data. Then, the Pearson correlation coefficient is used to quantify the strength and direction of linear relationships between variables. The computed Kaiser-Meyer-Olkin measure is 0.699, indicating high intercorrelations and substantial common variance among the variables, which suggests that the data is suitable for PCA. Bartlett's sphericity test yields a significant chi-square statistic with a p-value of 0, indicating that the dataset is suitable for PCA. Missing data is imputed using the mean, and all original data, which are positively oriented indicators, are normalized to prepare for subsequent analysis.

(2) Rotation of principal components: In order to enhance the interpretability and practical significance of the extracted principal components, rotation is performed to make the components more independent and facilitate a clearer interpretation of the data's variability. The author selects the common factors with eigenvalues greater than one and employs the Varimax rotation method to rotate the statistical values. The results of Varimax rotation are shown in Table 1.

Table 1. Results of Varimax rotation.

P.C.	Percentage of variance explained by eigenvalues			Sum of squared loadings (SSL)			Sum of squared loadings after rotation (SSLR)		
	Eig.	Var. (%)	Cum. (%)	Total	Var. (%)	Cum. (%)	Total	Var. (%)	Cum. (%)
1	3.322	48.68	48.68	3.322	48.68	48.68	3.295	48.05	48.05
2	1.517	27.25	75.93	1.517	27.25	75.93	1.528	26.97	75.02
3	1.001	11.51	87.44	1.001	11.51	87.44	1.087	12.42	87.44

Note: "P.C." represent principal components, "Eig." represents eigenvalue, "Var." represents variance, "Cum." represents cumulative.

In Table 1, it can be observed that the cumulative variance contribution of the three common factors is 87.44%, exceeding the explanation standard of over 80%. The component results of the three common factors after rotation are shown in Table 2.

Table 2. Component results of the three common factors after rotation (partial indicators).

P.C. Indicator	1	2	3
Annual per capita disposable income	0.156	0.854	0.156
Coverage rate of secure housing	0.698	0.125	0.198
Basic medical coverage rate	0.201	0.123	0.589

The first principal component shows a strong positive correlation with the coverage rate of secure housing, reflecting the basic living factors needed in poverty-stricken areas. These types of indicators are classified as social poverty reduction indicators. The second principal component exhibits a strong positive correlation with the annual per capita disposable income, which represents a policy poverty reduction indicator. These three principal components can explain a significant portion of the performance indicators of poverty alleviation projects.

(3) Component scores and ranking of poverty alleviation performance. With the aim of analyzing the performance level of poverty alleviation paired support projects, we start by obtaining the scores of the three principal components from the coefficient matrix. Then, we establish a comprehensive evaluation model using the variance contribution rates as weights, as shown in Table 3.

Table 3. Comprehensive evaluation model of poverty alleviation paired support project indicators (partial indicators).

P.C. Indicator	1	2	3
Annual per capita disposable income	0.384	0.054	0.156
Coverage rate of secure housing	0.099	0.485	0.198
Basic medical coverage rate	0.118	0.093	0.534

It can be seen that there is a linear relationship between the three principal components and the variable indicators. By using SPSS, the standardized values of each variable can be further calculated using the DEAP software to obtain the scores of each principal component for each year.

4. Problem Analysis and Recommendations

(1) Insufficient endogenous development momentum

Due to the small scale of village-level agricultural projects, lack of operational experience, and limited market demand for agricultural products, some industrial projects need more momentum for sustained development. As a result, they fail to generate significant economic benefits and have a limited impact on poverty reduction. Moreover, the remaining population in rural areas consists mainly of elderly, weak, sick, disabled, and young individuals, leading to a shortage of labor for industrial development. Data shows that by the end of 2020, the operating income of poverty-stricken households in the poverty alleviation villages supported by City A in City B accounted for only 5% of their total revenue. It is recommended to take further measures to enhance the endogenous development motivation of poor households.

(2) Insufficient investment in industries

Due to a significant amount of poverty alleviation funds being invested in industrial projects during the period of poverty alleviation, these projects are still generating benefits. However, the current focus on rural revitalization infrastructure construction still needs to be improved, resulting in a relatively low proportion of investment in

industrial projects. The proportion of investment in industrial projects is less than 15%. It is recommended to increase the proportion of investment in industries.

5. Conclusions

Based on DEA and PCA, we have revealed that the factors affecting the performance of poverty alleviation paired support projects from City A to City B include: (1) Insufficient support. This is mainly manifested in the shortage of resident village officials, significant differences in financial resources among different units, and inadequate division of corresponding support measures. (2) Weak foundation of poor households. This is mainly manifested in weak infrastructure, inability to scale up production, low levels of cultural and educational attainment among the population, severe aging, and the persistence of a “waiting for help” mentality among some poor households. (3) Limited investment in paired support. This is mainly manifested in limited government financial resources, difficulty in sustaining investment, and a low level of coordination and integration of financial resources for rural development.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

This work was supported by the Research Project of Guangzhou's Philosophical and Social Sciences Development “13th Five-Year Plan” in 2019, grant number 2019GZGJ125 and grant number 2019GZGJ124.

References

- [1] Wang, S.; Guo, Z. On China's precision poverty alleviation. *Guizhou Social Sciences*, 2015, 05, 147-150.
- [2] Shao, Y. “Third distribution” participates in helping Guangdong to achieve common prosperity. *Southern Daily*, 2022-10-05.
- [3] Lin, R. Performance evaluation research of poverty alleviation targeting in Danqing Town, Xiangxi Autonomous Prefecture. Master Thesis, Jishou University, 2022.
- [4] Zhao, H. Research on the performance of poverty alleviation in Shandong Province. Master Thesis, Wuhan Polytechnic University, 2016.
- [5] Zhang, Y. Evaluation and strategy research on the construction effectiveness of beautiful countryside from the perspective of rural revitalization strategy. Master Thesis, East China University of Science and Technology, 2019.
- [6] Kerzner, H. Project Management Metrics, KPIs, and Dashboards: A Guide to Measuring and Monitoring Project Performance. John Wiley & Sons, 2011.
- [7] Fleming, Q.; Koppelman, J. Earned value project management. Project Management Institute, 2010.
- [8] Atkinson, R. Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 1999, 17(6), 337-342.

- [9] Zhou, W.; You, J.; Zheng, H.A. Theoretical and methodological research on performance evaluation of government investment in science and technology projects. Chemical Industry Press, 2009.
- [10] Jiang, G. Government performance evaluation. Jinan University Press, 2016. (in Chinese)
- [11] Yang, Y.; Ma, B.; Koike, M. Efficiency-measuring DEA model for production system with k independent subsystems. *Journal of the Operations Research Society of Japan*, 2000, 43(3), 343-354.
- [12] Kruger, U.; Xie, L. Monitoring Multivariate Time-varying Processes. In *Statistical Monitoring of Complex Multivariate Processes*; Springer, 2012.
- [13] Anwar, A.; Olfati, M.; Pant, M.; et al. A Review on the 40 years of existence of data envelopment analysis models: Historic development and current trends. *Archives of Computational Methods in Engineering*, 2022, 29(7), 5397-5426.
- [14] Wang, M.Q.; Liang, L. Interval efficiency values and their ranking of decision-making units in CCR model. *Systems Engineering*, 2008, 26(4), 109-112.
- [15] Liu, J.; Lu, L.; Lu, W.; et al. A survey of DEA applications. *Omega*, 2013, 41(5), 893-902.



© 2022 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)