

Profitability Assessment Using Data Envelopment with Cluster Analysis: A Case for Different Types of Gas Stations

Ying Li^a, Shuhua Hou^b, Liming Yao^{*b}

^aBusiness School, Sichuan University, Chengdu 610064, P.R.China

^bUncertainty Decision-Making Laboratory, Sichuan University, Chengdu 610064, P.R.China
 lmyao@scu.edu.cn

The refined oil market in China is gradually opening up and becoming increasingly more competitive that urge the retail managers to assess the profitability of each gas station to increase its retail profits. This paper focuses on a profitability analysis of the gas station retail market by analyzing the sales and profitability which is commonly measured by examining the efficiency of the conversion of the inputs into outputs. Data envelopment analysis (DEA) is utilized to assess efficiency through calculating the relative efficiency of various gas stations. First of the analysis, testing for subsets of gas stations in order to implement classified management which has proven to be a good management practice in actual operation. This paper applies cluster analysis (CA) to increase the discriminatory power of DEA and improve the classification of gas stations. Then, 200 China National Petroleum Corporation (CNPC) gas stations in Sichuan province are analyzed to evaluate the applicability and effectiveness of the methodology. Next, the DEA results are compared to the SE-DEA results of traditional types of gas stations based on geographical environment. Then, several proposals are made for managers to develop retail marketing and expand refined products at terminal network stations.

1. Introduction

Recently the demand for gasoline in China has been rapidly increasing, and the focus on refined oil sales has shifted from the wholesale sector to the retail sector (Crompton et al., 2005). In this paper, we concentrate on the retail fuel segment, which has undergone a major transformation in recent years. Because the gasoline is always a hot topic, a large amount of valuable research has been conducted on the retail sector of gasoline stations. For example, (Shepard et al., 1991) investigated the control of gasoline prices by focusing on the retail price. (Kerzmann, 2014) studied a number of cost-based solutions for gasoline stations and examined the influence of shopping and search costs on pricing decisions. But few studies have focused on the profitability of the retail segment, which is a key measure of business operations. (Varga et al., 2014) summarizes the results of an engineering study to evaluate the techno-economic feasibility of different types of processes that producing high quality gas oil blending components from heavy gas oil fraction. (Capece et al., 2010) concentrate on the retail segment analyzing the profit and financial position of the companies and carried out in relation to the simultaneous effects of the various aspects of performance include profitability. In this study, we focus on the retail segment, which is more suitable for a comprehensive profitability analysis and its level of profitability directly determines the cost of a company and affects its development capacity and operational stability. Many evaluation models have been published in the literature on efficiency evaluation. (Song et al., 2014; Khayyam et al., 2015) put efforts on the energy efficiency of the transportation and production process. This paper concentrates on a profitability analysis of Sichuan branch sales at gas stations using data envelopment analysis, an effective tool for measuring the performance of the decision making units (DMU), which, in this study, are the terminal points of the gas station retail chain (Cook et al., 2010). As known as, DEA has been applied successfully to different entities that operate in various areas and in many contexts. Such as, (Tang et al., 2015) gave a detailed analysis of warehousing and distribution operations by DEA. (Katharina et al., 2011) argued that energy efficient manufacturing had been pushed to the top of the industry's agenda and derived requirements for energy management. (Yang et al., 2015) evaluated green

development efficiency of 31 municipalities and provinces in China. In this research, a CCR model (Cooper et al., 1978) is built to calculate the efficiency scores of the DMUs, and get the SE-DEA scores of traditional types of gas stations based on geographical environment through changing the CCR model (Yao et al., 2016). Due to the discriminatory power, this paper incorporates CA into the gas station profitability analysis to complement DEA by build an integrated approach of comprehensive profitability assessment. Before this study, (Samoilenko et al., 2008) has used cluster analysis for increasing the discriminatory power of DEA. (Ma et al., 2015) combines CA and DEA to evaluate the performance of renewable energy projects. (Dharmapala et al., 2015) strengthen the classification of profit-makers and do a comparative profitability study. The structure of the paper is as follows: Section 2 specifies the data set and presents the CA-DEA model; Section 3 illustrates the cluster composition reports and DEA results; Section 4 presents a contrastive analysis of the DEA results for different types of gas stations, identified by geographical environment and clusters based on the CA; Section 5 puts forward management proposal sand draws broad conclusions.

2. Methodology

Based on DEA and CA, this paper introduces an integrated approach to profitability assessment and optimization in gas station sectors. While it is always assumed that the DMUs in a sample have a functional similarity, the homogeneity of the DMUs are not checked, which severely limits the discriminatory power of the DEA results. In this section, the methodology is described. The general integrated model can be summarized as follows:

- I. Select representative indicators as DEA input and output indicators, as well as the CA grouping variables.
- II. Four clusters emerge from the CA and are then taken as the decision-making units for the DEA.
- III. Calculate the DEA scores and rank the DEA results. Compare the DEA results for the different types of gas stations identified by geographical environment which is common method adopted by gas station managers to classify or by the clusters based on the CA.
- VI. Finally, optimization suggestions for critical types of gas stations are proposed for each input sector.

2.1 Indicator selection

Considering many factors, the index system is eventually obtained which is related to evaluation and comprehensive information and accepted by managers of gas stations. The outputs represent the business performance and the inputs indicate the business' consumption of various economic resources. More specifically, the index system is shown in Table 1 and let X denote the input variables and Y denote the output variables. The indicators had a low level of inter-correlation with a correlation coefficient of less than 0.5.

Table 1: The input-output variables

Factor	Name of index	Symbol	Unit of index
Input	Fixed costs	X_1	RMB
	Variable expenses	X_2	RMB
	Total investment after conversion	X_3	RMB
Output	Sales	Y_1	t
	Profits	Y_2	RMB

2.2 CA-DEA model

In terms of the profitability analysis of gas stations, this paper identifies natural DMU groups based on their structural similarity with regards to levels of profitability. DMU structural similarity reflects the levels of the inputs and outputs that the DMUs receive and produce. Consequently, the first step in the methodology aims to determine four clusters, each of which have similar levels of received inputs and produced outputs. Among the various clustering methods, the technique employed by Ward was chosen as it generates a classification hierarchy while minimizing the variance within each group. Its conducted using SPSS (statistical package for the social sciences), which grouped the gas station samples in relation to the probability of certain values being scored by grouping variables. The grouping variables, which were also taken as the input and output indicators for the DEA, are introduced first. In this research, CCR model we used utilizes linear programming to determine the relative efficiencies of a set of homogeneous DMUs. This linear program calculates the performance scores by constraining all DMUs to an efficiency score of less than or equal to 1. According to the DEA efficiency definition, the relative efficiency of a DMU can be characterized as being either strong or weak. The DEA score for a decision unit of '1' indicates relatively good efficiency, meaning that the decision unit is integrated into a higher-output efficiency. That is, the DMU has reached maximum production efficiency compared with other DMUs. A score of less than one indicates relative inefficiency, meaning that it does not

use the given level of the inputs efficiently by producing a higher output level in the case of the input-oriented model.

Our study proposes a solution to the conduct of the DEA on a scaled heterogeneous data set using CA. This method identifies and takes into consideration the presence of heterogeneous subsets. Finally, we formulate an improved CA-DEA model to evaluate and optimize the profitability of gas stations according to cluster analysis (CA) and DEA theory. A chart of the evaluation model is shown in Figure 1.

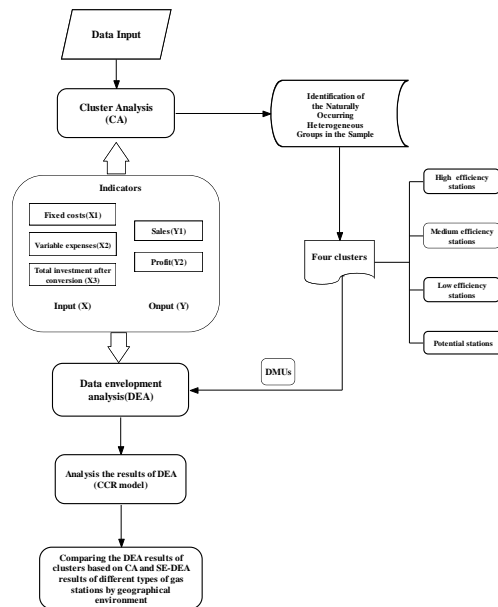


Figure 1: The framework of CA-DEA model

3. A case study

The sales branches of the China National Petroleum Corporation (CNPC) in Sichuan province made major adjustments to its oil sales strategy in 2009, proposing that it would increase retail accounting by more than 70% in the next few years. Further, CNPC requested its affiliated stations move their central focus from the wholesale to retail sectors. This section specifies the dataset and provides the results of the CA and DEA, then compares the DEA results from different types of gas stations that are identified by geographical environment or clusters based on the CA.

3.1 A dataset regarding CNPC gas stations in Sichuan province

The data for this study were obtained from the CNPC Sichuan petroleum sales firm. The program uses an equal probability sampling method for the overall sample, and a sample unit is a single gas station. After eliminating abnormal data, there were 197 samples collected, representing 16.68% of the total CNPC gas stations in Sichuan Province. The proportionate distribution of the samples in each region is shown in Figure 2.

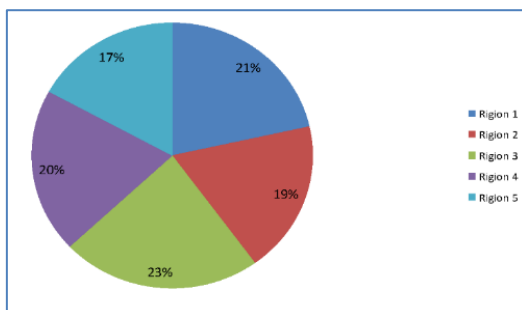


Figure 2: The proportionate distribution of the samples

3.2 Data envelopment analysis based on cluster analysis

In this section, the 2009 data were taken as the input and output data sources for the various stations being evaluated as shown in Table 3. From the input and output index difference analysis, it can be preliminarily concluded that the efficiency sorting was not consistent, which meant there was an efficiency deviation between the overall sample and the various gas stations. This indicates that profitability does not have a linear relationship with input-output efficiency. Hence, the DEA method was necessary to conduct a comprehensive profitability analysis for the various gas stations. The mean value of the four clusters drawn from the CA was collected as an evaluation data source, and a linear optimization model was built and solved using LINGO software. Taking DMU₁ as an example, the linear optimization model built is

$$\begin{aligned} & \min \theta \\ & \left\{ \begin{array}{l} 55.80\lambda_1 + 51.98\lambda_2 + 113.34\lambda_3 + 88.23\lambda_4 + 28.07\lambda_5 - 55.80\theta \leq 0 \\ 71.27\lambda_1 + 94.84\lambda_2 + 71.09\lambda_3 + 163.57\lambda_4 + 33.50\lambda_5 - 71.27\theta \leq 0 \\ 360.33\lambda_1 + 324.22\lambda_2 + 798.31\lambda_3 + 551.75\lambda_4 + 169.79\lambda_5 - 360.33\theta \leq 0 \\ 4159.24\lambda_1 + 5810.94\lambda_2 + 3577.88\lambda_3 + 12526.20\lambda_4 + 1468.27\lambda_5 - 4159.24 \geq 0 \\ 110.58\lambda_1 + 181.51\lambda_2 + 23.02\lambda_3 + 456.84\lambda_4 + 25.08\lambda_5 - 110.58 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ \lambda_j \geq 0, j = 1, \dots, 5 \end{array} \right. \end{aligned}$$

Table 2: Mean values of input-output indicators for overall sample and various gas stations

	Input indicators			Output indicators	
	Fixed costs (10 ⁴ RMB)	Variable expenses (10 ⁴ RMB)	TIAC (10 ⁴ RMB)	Sales (t)	Profit (10 ⁴ RMB)
Overall sample	55.80	71.27	360.33	4159.24	110.58
Cluster I	51.98	94.84	324.22	5810.94	181.51
Cluster II	113.34	71.09	798.31	3577.88	23.02
Cluster III	88.23	163.57	551.75	12526.20	456.84
Cluster IV	28.70	33.50	169.79	1468.27	25.08

NOTE: TIAC = Total investment after conversion

4. Results

The relative efficiency scores for the overall sample and the four clusters defined as the four types of stations based on the CA were determined, as shown in Table 3.

Table 3: Relative efficiency scores for the overall sample and the four clusters based on CA

Analysis unit	Overall sample	I	II	III
Scores of efficiency	0.762	0.800	0.657	1.000

According to the DMU effectiveness definition, only stations of Type III were considered DEA effective, and other DMUs were ineffective. Using the DEA scores in Table 3, the stations grouped into Type IV had the worst performance, and those grouped into Type III performed better. Type I stations had mediocre performance in terms of sales but had medium profitability. Type II stations had a high level of expenses and investments and, while the DMU showed a certain level of inefficiency, those stations demonstrated a large potential for better profitability. As a whole, the analysis indicated that the gas stations in Sichuan province were able to produce higher economic output with the current investment of resources. Type I, II and IV stations were medium efficiency stations, potential stations and low efficiency stations, respectively, as they did not achieve an optimal input-output ratio value in 2009. On the other hand, Type III were those which had the most appropriate investment scale according to the principle of minimum investment-maximum output. If investors put resources into such stations, they would achieve maximum benefits.

In this section, to verify the rationality and validity of the CA the DEA results for the various types of gas stations that were identified by geographical environment were compared to the clusters that were based on the CA. According to the regional market's geographical environment, the stations can be divided into four types: A, B, C, and D. "A stations" are located in the urban areas of prefecture-level cities. "B stations" are located on a highway or expressway in prefecture-level cities, connect suburbs with the transportation system.

“C stations” enjoy the largest development potential and space, and are located in county-level cities, towns. “D stations” are located in rural areas, towns, and on county roads with fewer changes in market environment. The mean value of the input-output indicators for the four types of gas stations are shown in Figure 3, which reveals no obvious difference between the input-output indicator means for A and B stations.

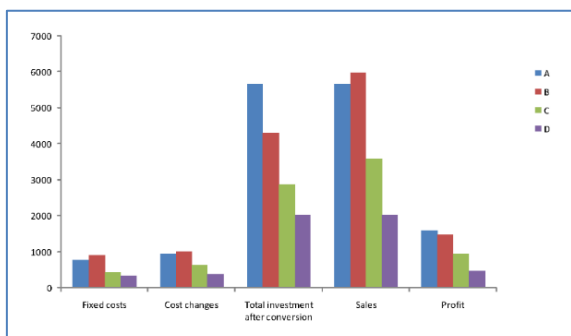


Figure 3: Mean values for the various gas stations classified by geographical environment

Because the DEA results using the CCR model mentioned in section 3.2 are all greater than 1, we removed from j -th linear combination of other inputs and outputs to get the scores for the SE-DEA which are listed in Table 4, showing that all are greater than 1. From the SE-DEA definition of DMU effectiveness, the four types of gas stations identified by geographical environment are all effective. From the regional market geographical environment, Type I gas stations suggest strong profitability for the individual indicators, oil profits per ton and profit levels. However, because Type I stations are located in urban areas of prefecture-level cities, the competition is more intense, and market growth tends towards saturation. As the city develops, land costs and other costs often increase considerably, space is constrained, and future profitability growth cannot exceed Types II or III. Therefore the future development potential of type I stations is small compared to other station categories.

Table 4: The SE-DEA scores for the various gas stations classified by geographical environment

DMU	Overall sample	I	II	III	IV
Scores of efficiency	1.1023	1.0953	1.1229	1.1268	1.0992

The results of SE-DEA indicate that all stations in Sichuan province are efficient, and they all achieve optimal profitability. However, there were some gas stations that indicated poorer operating situations. Due to the method of classification, the DMU means used in the DEA did not provide accurate indicator levels for most of sample sites in each type.

5. Proposals and Conclusions

From the view of the development potential of gas stations, a company's key strategic layout can be adjusted according to the development potential of all types of gas stations in the future. According to the sorted DEA efficiency evaluation of the various types of gas stations, in the future, the company should focus on the development of stations Types II and III stations, should consolidate the development of Type I and take into account the development of Type IV.

In conclusion, A CA-DEA model was developed in this paper to accurately evaluate the profitability efficiency of different types of gas stations. Because the gas station sample included stations with different levels of development and profitability, two or more subsets were identified, which constrained the discriminatory power of the DEA. To address this problem, a cluster analysis was conducted to determine the gas station classifications. In the paper, the CA-DEA model was applied to a profitability evaluation of gas stations in Sichuan province. Next, the comprehensive profitability was analyzed within each cluster. The relative efficiencies of each subset demonstrated that only one type of station was efficient. To illustrate the effectiveness and rationality of the CA-DEA model, the DEA results for the various types of gas stations, which were identified by geographical environment, were compared to the clusters based on the CA. From the result, some proposals were made to assist decision-makers in improving profitability efficiency and to offer investors an approach for gas station classifications and profitability evaluation. This model is useful not only for the oil market retail sector but also for other areas, such as the new energy market, fast moving consumer goods (FMCG) and other businesses that require efficiency evaluations. There are, however, improvements

that could be made to improve the discriminatory power of DEA results and suitable CA conducted method for specific issues.

Acknowledgments

This research was supported by the National Natural Science Foundation for Young Scholars of China (Grant No. 71301109), the Western and Frontier Region Project of Humanity and Social Sciences Research, Ministry of Education of China (Grant No. 13XJC630018) and the Central Universities Fundamental Research Project (Grant No. skqy201653).

Reference

- Borenstein S., 1991. Selling costs and switching costs: explaining retail gasoline margins. *The RAND Journal of Economics*, 354-369.
- Bunse K., Vodicka M., Schönsleben P., Brühlhart M., Ernst F.O., 2011. Integrating energy efficiency performance in production management-gap analysis between industrial needs and scientific literature. *Journal of Cleaner Production*, 19(6), 667-679, DOI: 10.3390/su7044492
- Capece G., Cricelli L., Pillo F. D., Leviardi N., 2010. A cluster analysis study based on profitability and financial indicators in the Italian gas retail market. *Energy Policy*, 38(7), 3394-3402, DOI: 10.1016/j.enpol.2010.02.013
- Charnes A., Cooper W.W., Rhodes E., 1978. Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444. DOI:10.1016/0377-2217(78)90138-8
- Cook W.D., Liang L., Zhu J., 2010. Measuring performance of two-stage network structures by DEA: a review and future perspective. *Omega*, 38(6), 423-430. DOI: 10.1016/j.omega.2009.12.001
- Crompton P., Wu Y., 2005. Energy consumption in China: past trends and future directions. *Energy economics*, 27(1), 195-208. DOI: 10.1016/j.eneco.2004.10.006
- Dharmapala P. S., Edirisuriya, P., 2015. Classification of profitability using DEA and cluster analysis: a transnational comparison of South Asian banks. *International Journal of Information and Decision Sciences*, 7(3), 213-240. DOI: 10.1504/IJIDS.2015.071370
- Kerzmann T.L., Buxton G.A., Preisser J., 2014. A computer model for optimizing the location of natural gas fuelling stations. *Sustainable Energy Technologies and Assessments*, 7, 221-226. DOI: 10.1016/j.seta.2013.10.004
- Khayyam H., Naebe M., Bab-Hadiashar A., Jamshidi F., Li Q., Atkiss S., Fox B., 2015. Stochastic optimization models for energy management in carbonization process of carbon fibre production. *Applied Energy*, 158, 643-655, DOI: 10.1016/j.apenergy.2015.08.008
- Ma X., Zeng B., Zhang Y., Li Y., & Liu Z. (2015, November). Comprehensive evaluation of renewable energy for power projects based on CA-DEA model. In 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT) (1848-1853). IEEE.
- Samoilenko S., Osei-Bryson K. M., 2008. Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. *Expert Systems with Applications*, 34(2), 1568-1581, DOI: 10.1016/j.eswa.2007.01.039
- Song M., Wu N., Wu K., 2014. Energy consumption and energy efficiency of the transportation sector in Shanghai. *Sustainability*, 6(2), 702-717, DOI: 10.3390/su6020702
- Tang L., Peng Y., Xiao Z., 2015. Analysis and evaluation of relative efficiency of warehousing and distribution operations based on mixed DEA model. *Chemical Engineering Transactions*, 46, 583-588, DOI: 10.3303/CET1546098
- Varga Z., Eller Z., Hancsók J., 2014. Techno-economic evaluation of quality improvement of heavy gas oil with different processes. *Chemical Engineering Transactions*, 39, 1657-1662. DOI: 10.3303/CET1439277
- Yang Q., Wan X., Ma H., 2015. Assessing Green Development Efficiency of Municipalities and Provinces in China Integrating Models of Super-Efficiency DEA and Malmquist Index. *Sustainability*, 7(4), 4492-4510
- Yao Y. Y., & Zhang R. S., 2016. Empirical Research on Efficiency Measure of Financial Investment in Education Based on SE-DEA. In *Fuzzy Systems & Operations Research and Management* (389-402). Springer International Publishing.