Expert Systems with Applications 39 (2012) 1118–1123

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Input/output indicator selection for DEA efficiency evaluation: An empirical study of Chinese commercial banks

Yan Luo *, Gongbing Bi, Liang Liang

School of Management, University of Science and Technology of China, Hefei, Anhui Province 230026, PR China

ARTICLE INFO

Keywords:
Data envelopment analysis (DEA)
Variables selection
Classification
Cash value added
Regression

ABSTRACT

One of the interesting research subjects in DEA is to choose appropriate input and output indicators. In the process, one may encounter many problems, such as the selection tools, correlation analysis, and the classification of input versus output status. In this paper, we propose a new method for choosing DEA variables. Unlike previous research, it is based on the conception of cash value added (*CVA*), and can make a selection according to the statistic results. This new method has some advantages: first, it is more objective, avoiding the influence of subjective factors on the subsequent calculation; second and most important is that it provides managers and researchers with measurement variables and exact classifications of these factors; third, all variables under discussion come from financial statements which are easily available. This variable selection method has been applied to 14 Chinese commercial banks, and both regression and statistic test results are satisfactory.

1. Introduction

Data envelopment analysis (DEA) is a non-parametric technique for measuring the relative efficiency of a set of similar units, usually referred to as DMUs which convert multiple inputs to multiple outputs. Since the advent of DEA (Charnes, Cooper, & Rhodes, 1978), this methodology has been applied in a wide range of applications, such as schools, banks, and hospitals. A substantial amount of scholarly effort has been devoted to its development over the past three decades (Cook & Seiford, 2009).

In the seminal work of Charnes et al., efficiency is represented by the ratio of weighted outputs to weighted inputs. Some expanded forms were proposed by other researchers later. Central to each formulation is the need to account for variables' selection. Just as Paradi, Vela, and Yang (2004), pointed out, besides the choice of DEA technology (model), selecting inputs and outputs is the other crucial consideration that the analyst must keep in mind. Unfortunately, there has been inadequate attention to this issue, while most attention has been given to building models. Several researchers have just chosen variables subjectively or calculated efficiency based directly on others' selection results. Lack of publicly available data is perhaps one of the reasons behind this trend.

In some cases, there are too many variables, which violate the rule of the thumb (Cooper, Seiford, & Zhu, 2004) and causes bad discrimination between DMUs. Actually, certain factors may be eliminated because some variables prove to be very highly correlated. Several methods can be used to address this problem. Principal component analysis (PCA) is one of the most common techniques, which is widely used in multivariate statistics (Luukka, 2009; Ramalingam & Charles, 2007; Wang & Du, 2000). Azadeh and Ghaderi (2007) integrated DEA and PCA for quality assessment of products. However, PCA just provides a way to reduce the number of inputs and outputs when there is a large-dimension data set after the choice of variables. There is also a graphic display method, called Co-Plot, which can be used to run a correlation analysis on all variables and DEA results (Raveh, 2000). But before calculation, one must decide the classification of every variable, because the process begins with the ratio of output/input (Adler & Raveh, 2008). In fact, the above two methods cannot solve the problem with variables selection.

As is known, the usual variables in DEA are such that more is better for output, and less is better for input. However, the behaviors of many variables are just opposite to this in some situations, such as air pollution. Traditional DEA may cause confusion when dealing with such problems. Several researchers solve this problem by making restrictions to specific variables (Fare & Grosskopf, 2004; Hua & Bin, 2008). However, these ideas are more related to defining models rather than variables selection, and how to deal with such variables depends on what the model is intended to achieve.

In the traditional application of DEA, it is assumed that the status of variables is provided with a priori. However, in the real world, there exists a kind of variable characterized as "flexible", i.e., a variable that acts as an input and an output at the same

* Corresponding author. Tel.: +86 551 3603461; fax: +86 551 3600025.
E-mail address: yanluo@mail.ustc.edu.cn (Y. Luo).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.07.111

time. Beasley (1995) first discussed this issue and presented a formulation for this situation in evaluating universities in the UK. Cook, Green, and Zhu (2006) later gave an alternative version of Beasley's. No matter what the models are, they should not consider the influence of such variables in both places, but rather consider it in the most appropriate place.

In this paper, we propose a new methodology for choosing variables and making decisions on which variables to include as inputs and which ones as outputs at the same time. The new method is based on the concept of cash value added (*CVA*), which was first introduced by Erik Ottosson and Fredik Werssenrieder. In the theory of modern business management, cash flow plays a more and more important role in an enterprise's development. It reflects the state of a business most accurately and decides the growth and decline of the enterprise in a certain degree. Many companies go bankrupt due to the deterioration of their cash flow. Unfortunately, the financial crisis that broke out in 2008 has made it hard for many enterprises to operate, and cash flow risks have gradually emerged. Under this situation, we consider cash flow in the DEA framework for efficiency evaluation. Appropriate variables are selected by our method in view of their influence on the DMUs' cash flow. A factor can be taken as an output if it has a positive influence on a DMU's cash flow or as an input otherwise. Finally, statistic tests are applied to analyze the selection results, demonstrating the validity of our method.

What distinguish this research from those exists in the literature above mainly three issues. First, it is more objective. We are to control the influence of subjective factors and let the data speak for itself. Second, also the most important, it provides the managers and researchers with relevant variables and exact status designations of them. Third, the data can be easily obtained since all variables concerned are from balance sheet and statement of cash flows.

The rest of the paper is arranged as follows. Section 2 gives a simple introduction to *CVA* and discusses how variables are selected. In Section 3, an example of 14 Chinese commercial banks is presented to illustrate the proposed method. Conclusions are drawn in Section 4.

2. *CVA* and variable selection in DEA

2.1. CVA

Nowadays, more and more corporations spend a large amount of time and money on evaluating their operational performance. Traditional performance evaluation always focuses on the financial aspect; however, financial data about revenue and profit may be untrue because of creative accounting. As a result, it is often difficult to appraise a business's current situation objectively and accurately, and hard to correctly evaluate the operational performance of an enterprise and its abilities of creating values.

Cash value added (*CVA*), representing the value creation from the shareholders' point of view, is a new framework for value appraisal based on the integration of the operational capital, risk management of capital cost and operation of investment program. Unlike the traditional method, it utilizes data from the statement of cash flows which is prepared on the cash basis of accounting. The data in the statement of cash flows usually need the certification of specific organizations, such as banks, to prevent the influence of creative accounting. At the same time, cash flow does not affect items that investors are mostly concerned with, such as revenue and profit. Thus, it is a useful tool to objectively evaluate corporate performance, and provide meaningful information for managers' investment decision.

CVA is a net present value model and classifies investments into two categories, strategic and non-strategic investments. The former, which form the capital base in the *CVA* model, are used to create new value for shareholders. The latter are considered as "cost", with the purpose of maintaining the value created by the former. The *CVA* model comes down to operating cash flow demand (*OCFD*) and operating cash flow (*OCF*). *OCFD* is a serial of the cash flow, which is the sum of discounted strategic investment in every period, i.e., a net present value over the strategic investment's economic life. *OCF* is the cash flow before strategic investments but after non-strategic investments, and helps the strategic investment to create value. *CVA* can be calculated by the follow formulation:

$$CVA = OCF - OCFD \tag{1}$$

Obviously, a strategic investment creates value if the *OCF* exceeds the *OCFD* over time.

2.2. Variable selection methodology

There are two crucial considerations in DEA: the choice of DEA technology (or DEA model) and the variable selection. In the past three decades, several researchers have made much effort in these areas. However, many problems still exist in research, such as the choice of selection tools, correlation analysis on variables, and the classifications of input versus output status when selecting variables.

To solve such problems, we propose a new method for choosing variables. Our selection approach is built on but different from *CVA*. *CVA* is based solely on cash flow but our method is based on both the cash flow and balance sheet. The other difference is that *CVA* is a tool to evaluate an investment decision, while ours is a more comprehensive method covering the whole operating process. As mentioned above, cash flow is more persuasive than the traditional profit index in explaining a company's performance. We take this idea into the DEA application, i. e., choose appropriate input and output variables according to their influence on the DMU's cash flow.

Suppose there are n DMUs. Every DMU has the same type of measure index $X = (x_1, x_2, \cdots, x_k)$, where $x_i (i = 1, \ldots, k)$ is the item selected from the balance sheet. The cash flow of every DMU is denoted as *CF* for abbreviation. Our goal is to select DEA input and output variables from X and offer a clear status designation for each one.

The proposed procedure that serves our purposes is discussed below.

Step 1: Some representative items $x_i (i = 1, \ldots, k)$ are selected from the balance sheet. The factors with low numerical values and having little influence on DMU's operation, and those whose numerical value change severely are out of our consideration. In this way, we select a majority of items in the balance sheet except some particular ones so as to avoid subjectivity.

Step 2: *CF* selected from the statement of cash flows is treated as a dependent variable, and every x_i selected in step one as an independent variable. One should keep in mind, however, that *CF* is determined by the variations of those independent variables. Using statistic regression analysis, we can find factors that influence DMU's *CF* and value increment further. In this process, the coefficients of all factors are derived from the solution of the regression analysis. Then, the classification of inputs and outputs will be decided in light of the signs of these coefficients.

a. If an independent variable is positively correlated with *CF*, i.e., its coefficient is a positive number, then, it is deemed as an output.