Available online at www.sciencedirect.com

ScienceDirect

European Journal of Operational Research 196 (2009) 312–322

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

O.R. Applications

Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks

Chiang Kao [a,*], Shiang-Tai Liu [b]

[a] Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan, ROC
[b] Graduate School of Business and Management, Vanung University, Chung-Li, Tao-Yuan 320, Taiwan, ROC

Received 25 June 2006; accepted 14 February 2008
Available online 26 February 2008

Abstract

Conventional data envelopment analysis (DEA) for measuring the efficiency of a set of decision making units (DMUs) requires the input/output data to be constant. In reality, however, many observations are stochastic in nature; consequently, the resulting efficiencies are stochastic as well. This paper discusses how to obtain the efficiency distribution of each DMU via a simulation technique. The case of Taiwan commercial banks shows that, firstly, the number of replications in simulation analysis has little effect on the estimation of efficiency means, yet 1000 replications are recommended to produce reliable efficiency means and 2000 replications for a good estimation of the efficiency distributions. Secondly, the conventional way of using average data to represent stochastic variables results in efficiency scores which are different from the mean efficiencies of the presumably true efficiency distributions estimated from simulation. Thirdly, the interval-data approach produces true efficiency intervals yet the intervals are too wide to provide valuable information. In conclusion, when multiple observations are available for each DMU, the stochastic-data approach produces more reliable and informative results than the average-data and interval-data approaches do.

Keywords: Data envelopment analysis; Efficiency; Stochastic data; Interval data

1. Introduction

Since the paper of Charnes et al. (1978), data envelopment analysis (DEA) has been widely discussed from the methodological as well as practical side in measuring the relative efficiency of not-for-profit organizations that utilize the same inputs to produce the same outputs (see the review of Lovell, 1994; Seiford, 1996). Even for profit-driven organizations such as banks, manufacturing firms, hotels, etc., applications have also been reported. DEA is now a standard technique for performance evaluation for all types of decision making units (DMUs).

As indicated by its name, DEA is based on observed data. It is implicitly assumed that each DMU to be evaluated only has one set of deterministic data. If more than one set of data appears, they are usually averaged to result in one set. A more delicate way is to perform a time-series type analysis, e.g., window analysis (Charnes et al., 1985), to investigate the stability of the result. The efficiency measure is still a point value, without any intervals or distributions associated with it.

Under the assumption of variable returns-to-scale, the DEA model for measuring the efficiency of the kth DMU, E_k, in a set of n DMUs can be formulated as follows (Banker et al., 1984):

$$\begin{aligned} E_k = \max \ & \sum_{r=1}^{s} u_r Y_{rk}, \\ \text{s.t.} \quad & v_0 + \sum_{i=1}^{m} v_i X_{ik} = 1, \\ & \sum_{r=1}^{s} u_r Y_{rj} - \left(v_0 + \sum_{i=1}^{m} v_i X_{ij}\right) \leqslant 0, \ j = 1, \ldots, n, \\ & u_r, v_i \geqslant \varepsilon, \ r = 1, \ldots, s; \ i = 1, \ldots, m, \end{aligned} \tag{1}$$

* Corresponding author. Tel.: +886 6 2753396.
E-mail addresses: ckao@mail.ncku.edu.tw (C. Kao), stliu@vnu.edu.tw (S.-T. Liu).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.02.023

where X_{ij} and Y_{rj} are the ith, $i = 1, \ldots, m$, input and rth, $r = 1, \ldots, s$, output, respectively, of the jth, $j = 1, \ldots, n$, DMU, and ε is a small non-Archimedean number (Charnes and Cooper, 1984) for restricting the DMU to assign 0 weight to unfavorable factors. In practice, the data collected are often snapshot observations, this is true even for annual data, only the exposure time lasts for one year. If one pressed the shutter a few seconds later, one would probably get different observations. This implies that the observations are stochastic, and so the associated efficiency measures should be stochastic as well.

Many articles have addressed the issue of stochastic observations. Most of them concentrate on the case of single-output simply because it is easier to manipulate. The discussion of stochastic observation starts with the parametric approach where the functional form of the production frontier needs to be specified beforehand. Aigner et al. (1977) and Meeusen and van den Broeck (1977) are the pioneers of work on stochastic frontier. Efficiency is a measure of deviation of the actual position of the DMU from the production frontier. The stochastic frontier approach takes the measurement error and statistical noise into account, and allows the separation of the deviation into inefficiency and noise components. This feature has attracted many theoretical studies as well as practical works following this approach (Bauer, 1990; Greene, 1993). A tradeoff in gaining the advantage in decomposition is the effort spent in acquiring the functional and distributional form for production and error components, respectively. Mistakenly specified functions lead to erroneous efficiency measures, and its regression-type analysis only allows for one output.

DEA is a nonparametric approach, in that the functional form of the production frontier is not required. In its early stage of development, the DEA approach has been criticized for its lack of statistical properties (Greene, 1993; Schmidt, 1985). Banker (1993) lays the statistical foundation for DEA by showing that the DEA estimators are also maximum likelihood estimators under certain conditions. Gijbels et al. (1999) derive the asymptotic distribution of the DEA estimator for the single-input and single-output case. Kneip et al. (1998) generalize the results to the multi-input and multi-output case. Simar and Wilson (1998, 2000) propose a bootstrap strategy to analyze the sensitivity of efficiency scores relative to the sampling variations of the estimated frontier. Some scholars believe the deterministic approach is conceptually flawed because it does not allow measurement errors. Interestingly, several simulation studies (Banker et al., 1993; Ondrich and Ruggiero, 2001; Ruggiero, 1999, 2004) show that the stochastic approach does not accurately decompose the total error into inefficiency and noise components. At best, the stochastic frontier is only as good as the deterministic model. The former is superior only in cases when the assumed technology is close to the given underlying technology.

One approach for dealing with stochastic data in DEA is chance constrained programming. Due to the stochastic nature of the observations, the constraints that require the aggregated output to be smaller than the aggregated input are to be satisfied with specified probabilities. Different probability determines different efficiencies for the set of DMUs (Cooper et al., 2002, 1996; Land et al., 1994; Olesen and Petersen, 1995). In this approach, it is assumed that the efficiency of a DMU is stochastic, and the observation is just an occurrence of random phenomenon.

Another approach is to treat the uncertain observation as interval data and calculate interval efficiency (Cooper et al., 1999, 2001; Despotis and Smirlis, 2002; Kao, 2006). Although interval efficiency provides more information regarding the range of the stochastic efficiency, the distribution or even the mean of efficiency is still not known. Apparently, if the intervals are narrow, then the resulted efficiency intervals will also be narrow, and they will be representative. If, on the contrary, the intervals are rather wide, then the resulted efficiency intervals will be wide as well. In this case, not much information is provided, and the base for making decision is consequently weak. The case of the commercial banks in Taiwan is an example, where the observations of some input/output factors have wide variation at different years.

In this paper, we will show that the conventional way of using the average data of several years to calculate the efficiency leads to erroneous efficiency measures. Treating the observations as interval data obtains efficiency intervals which are too wide to draw conclusions. To avoid these drawbacks, this paper treats the data as stochastic and finds the distributions of the input/output data of each bank. By applying a simulation technique, the efficiency distribution of each bank is obtained. Efficiency distributions are obviously more informative than either efficiency scores or efficiency intervals for drawing appropriate conclusions.

This paper is organized as follows. Firstly, we use a simple example to explain the concept of distributional efficiency in stochastic DEA. A graphical expression of the production frontier illustrates the difference between using the average data and stochastic data to measure the relative efficiency of each DMU. Then we describe the stochastic nature of the Taiwan commercial banks, and apply a simulation technique to find the efficiency distribution of each bank. Finally, the accuracy in approximating the distribution is discussed and the results of average data, interval data, and stochastic data are compared. Some conclusions are drawn from the discussions and comparisons.

2. Graphical illustration

For simple problems, the efficiency distribution of each DMU can be derived analytically. In Fig. 1, there are five DMUs, labeled as A, B, C, D, and E, using input X at levels 1, 2, 3, 4, and 5, respectively, to produce output Y of the amounts of 2, 5, 5, 6, and 5. Under the assumption of variable returns-to-scale, the production frontier constructed from these five DMUs is the piecewise line segments $ABDD'$, where A, B, and D are on the frontier and are thus