European Journal of Operational Research 239 (2014) 1–16

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Invited Review

Network data envelopment analysis: A review

Chiang Kao *

Department of Industrial and Information Management, National Cheng Kung University, Tainan, Taiwan, Republic of China

ARTICLE INFO

Article history:
Received 3 November 2013
Accepted 18 February 2014
Available online 24 February 2014

Keywords:
Data envelopment analysis
Network
Efficiency

ABSTRACT

Network data envelopment analysis (DEA) concerns using the DEA technique to measure the relative efficiency of a system, taking into account its internal structure. The results are more meaningful and informative than those obtained from the conventional black-box approach, where the operations of the component processes are ignored. This paper reviews studies on network DEA by examining the models used and the structures of the network system of the problem being studied. This review highlights some directions for future studies from the methodological point of view, and is inspirational for exploring new areas of application from the empirical point of view.

1. Introduction

Efficiency measurement is an important task in management, to better understand the past accomplishments of a unit and planning for its future development. Since the seminal work of Charnes, Cooper, and Rhodes (1978), data envelopment analysis (DEA) has been widely recognized as an effective technique for measuring the relative efficiency of a set of decision making units (DMUs) that apply multiple inputs to produce multiple outputs, with many theoretical developments and practical applications being reported (see, for example, the review of Cook & Seiford, 2009; Emrouznejad, Parker, & Tavares, 2008; Liu, Lu, Lu, & Lin, 2013a, 2013b; Seiford, 1996; Zhou, Ang, & Poh, 2008).

DEA was originally developed to measure the efficiency of a DMU as a whole unit, without considering its internal structure. In other words, the system is treated as a black box, within which inputs are supplied to produce outputs, with there generally being a positive correlation between the two. However, there are empirical studies that indicate this may not always be true. For example, Cron and Sobol (1983) showed that IT (information technology) had little effect on business performance. A later study found that the operation of banking and similar industries had two processes, capital collection and investment (Wang, Gopal, & Zionts, 1997), and that while IT was useful for the former, whether the firms would actually make a profit or not was dependent on correct investment decisions being made. This indicates that to study the performance of a DMU, it is necessary to study its component processes, so that the cause of any inefficiencies can be identified.

The first paper discussing this idea is probably Charnes et al. (1986), which found that army recruitment had two processes: the first created awareness through advertisements, and the second created contracts. Separating large operations into detailed processes helps identify the real impact of input factors. The simplest case is to separate the whole operation into two processes, as Charnes et al. (1986) and Wang et al. (1997) did. There are many more complicated cases in which the whole operation is separated into more than two processes. These may have a series structure, a parallel structure, or a mixture of these. These structures are generally called network structures, and the DEA technique to measure the efficiency of systems with a network structure is called network DEA (Färe & Grosskopf, 2000).

Another issue that should be noted is that ignoring the operations of the component processes may obtain misleading results, and a number of examples have been presented to show that an overall system may be efficient, even while all component processes are not (Kao & Hwang, 2008). More significantly, there are cases in which all the component processes of a DMU have performances that are worse than those of another DMU, and yet the former still has the better system performance (Kao & Hwang, 2010). These findings indicate that a network DEA model is required to produce correct results when measuring efficiencies.

Hundreds of works that discuss network DEA have been published since Charnes et al. (1986). Some develop models to measure efficiencies under specified conditions, some examine the properties possessed by certain models, and others apply existing models to solve real world problems. Cook, Liang, and Zhu (2010) reviewed a number of models for the basic two-stage system, in which the system has only two processes connected in series, and the second only consumes all the outputs from the first for production. Castelli, Pesenti, and Ukovich (2010) reviewed

* Tel.: +886 6 2753396.
E-mail address: ckao@mail.ncku.edu.tw

http://dx.doi.org/10.1016/j.ejor.2014.02.039
0377-2217/

shared-flow, multilevel, and some network models. The network models they reviewed are of the general network DEA form developed by Färe and Grosskopf (2000), leaving many others untouched.

This paper carries out a systematic review of studies related to network DEA. The classification of the articles is based on the model developed or used in the study, and the structure of the network systems of the problems studied. For each type of structure, the models, applications, and data types that have appeared in the literature are reviewed. It is anticipated that the results of the current study will inspire the development of new models, the solving of practical problems, and the presentation of new applications in network DEA.

2. Efficiency measurement

Let X_{ij} and Y_{rj} denote the ith input ($i = 1,\ldots,m$) and rth output ($r = 1,\ldots,s$) of the jth DMU ($j = 1,\ldots,n$). The DEA model developed by Charnes et al. (1978) for measuring the relative efficiency of DMU 0 under the assumption of constant returns to scale in multiplier form is as follows:

$$\begin{aligned} E_0 = \max.\quad & \sum_{r=1}^{s} u_r Y_{r0} \\ \text{s.t.}\quad & \sum_{i=1}^{m} v_i X_{i0} = 1 \\ & \sum_{r=1}^{s} u_r Y_{rj} - \sum_{i=1}^{m} v_i X_{ij} \leqslant 0, \quad j = 1,\ldots,n \\ & u_r, v_i \geqslant \varepsilon, \quad r = 1,\ldots,s, \quad i = 1,\ldots,m \end{aligned} \tag{1}$$

where u_r and v_i are virtual multipliers, and ε is a small non-Archimedean number used to avoid ignoring any factor in calculating efficiency (Charnes & Cooper, 1984). This model is conventionally referred to as the CCR model. If the returns to scale are allowed to be variable, then an unrestricted variable u_0 is subtracted from $\sum_{r=1}^{s} u_r Y_{r0}$ in the objective function and $\sum_{r=1}^{s} u_r Y_{rj}$ in the constraint set (Banker, Charnes, & Cooper, 1984). Model (1) is input-oriented. The DEA model can also be formulated as an output-oriented one. In this case, the model under constant returns to scale is the same as Model (1), whereas the one under variable returns to scale adds an unrestricted variable v_0 to $\sum_{i=1}^{m} v_i X_{i0}$ and $\sum_{i=1}^{m} v_i X_{ij}$ in the constraint set (Banker et al., 1984). The model which allows returns to scale to be variable is usually referred to as the BCC model.

Model (1) has a dual, which can be formulated as follows:

$$\begin{aligned} E_0 = \min.\quad & \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j X_{ij} + s_i^- = \theta X_{i0}, \quad i = 1,\ldots,m \\ & \sum_{j=1}^{n} \lambda_j Y_{rj} - s_r^+ = Y_{r0}, \quad r = 1,\ldots,s \\ & \lambda_j, s_i^-, s_r^+ \geqslant 0, \quad j = 1,\ldots,n, i = 1,\ldots,m, r = 1,\ldots,s \\ & \theta \text{ unrestricted in sign} \end{aligned} \tag{2}$$

This model is input-oriented, and is of the envelopment form. If a model with an output orientation is desired, then the objective function is changed to "max. $\theta + \varepsilon(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+)$" and the variable θ attached to X_{i0} is moved to Y_{r0}. Furthermore, when the assumption of constant returns to scale is changed to variable returns to scale, then a convexity constraint of $\sum_{j=1}^{n} \lambda_j = 1$ is added.

Model (1) (or Model (2), equivalently) does not take the internal structure of the system into account in measuring efficiency, and thus is usually called the black-box model. The black-box model only considers the inputs X_i consumed by and the outputs Y_r produced from the system. In contrast to the black-box model, a network model takes the operations of the component processes into account in measuring efficiency. When the internal structure of a system is considered, the inputs supplied from outside can be used directly by all processes, and the outputs of every process can be either the final outputs of the system or the intermediate products to be used by other processes for production.

Suppose a system is composed of p processes. Other terms, such as subunits, sub-DMUs, divisions, and components, have also been used, and this paper will use the term processes when there is no ambiguity. Denote $X_{ij}^{(k)}$ and $Y_{rj}^{(k)}$ as the ith input supplied from outside, $i \in I^{(k)}$, where $I^{(k)}$ is the index set of the exogenous inputs used by process k, and the rth final output of the system, $r \in O^{(k)}$, where $O^{(k)}$ is the index set of the final outputs produced by process k, $k = 1,\ldots,p$, respectively, of the jth DMU. Clearly, the sums of $X_{ij}^{(k)}$ and $Y_{rj}^{(k)}$ for all p processes are the system input X_{ij} and system output Y_{rj}, respectively, i.e., $\sum_{k=1}^{p} X_{ij}^{(k)} = X_{ij}$ and $\sum_{k=1}^{p} Y_{rj}^{(k)} = Y_{rj}$. Further, let $Z_{fj}^{(a,k)}$ denote the fth intermediate product produced by process a, $f \in M^{(k)}$, where $M^{(k)}$ is the index set of the intermediate products used by process k, and $Z_{gj}^{(k,b)}$ denote the gth intermediate product to be used by process b, $g \in N^{(k)}$, where $N^{(k)}$ is the index set of the intermediate products produced by process k. The same intermediate product f produced by different processes for process k to use can be aggregated as $Z_{fj}^{(k)}$, i.e., $\sum_{a=1}^{p} Z_{fj}^{(a,k)} = Z_{fj}^{(k)}$. Similarly, the same intermediate product g produced by process k for different processes to use can also be aggregated as $Z_{gj}^{(k)} = \sum_{b=1}^{p} Z_{gj}^{(k,b)}$. With these notations, the general network structure can be depicted, as shown in Fig. 1.

Assuming the most general case where the technologies of all processes are allowed to be different, the production possibility set defined by the general network structure of Fig. 1 is $T = \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) | \sum_{j=1}^{n} \lambda_j^{(k)} X_{ij}^{(k)} \leqslant x_i,\ i \in I^{(k)},\ \sum_{j=1}^{n} \lambda_j^{(k)} Y_{rj}^{(k)} \geqslant y_r,\ r \in O^{(k)},\ \sum_{j=1}^{n} \lambda_j^{(k)} Z_{fj}^{(k)} \leqslant z_f,\ f \in M^{(k)},\ \sum_{j=1}^{n} \lambda_j^{(k)} Z_{gj}^{(k)} \geqslant z_g,\ g \in N^{(k)},\ \lambda_j^{(k)} \geqslant 0,\ j = 1,\ldots,n,\ k = 1,\ldots,p\}$. The input-oriented model for measuring the system efficiency can be formulated as:

$$\begin{aligned} \min.\quad & \theta \\ \text{s.t.}\quad & \sum_{j=1}^{n} \lambda_j^{(k)} X_{ij}^{(k)} + s_i^{-(k)} = \theta X_{i0}^{(k)}, \quad i \in I^{(k)}, \quad k = 1,\ldots,p \\ & \sum_{j=1}^{n} \lambda_j^{(k)} Y_{rj}^{(k)} - s_r^{+(k)} = Y_{r0}^{(k)}, \quad r \in O^{(k)}, \quad k = 1,\ldots,p \\ & \sum_{j=1}^{n} \lambda_j^{(k)} Z_{fj}^{(k)} + s_f^{o(k)} = Z_{f0}^{(k)}, \quad f \in M^{(k)}, \quad k = 1,\ldots,p \\ & \sum_{j=1}^{n} \lambda_j^{(k)} Z_{gj}^{(k)} - s_g^{o(k)} = Z_{g0}^{(k)}, \quad g \in N^{(k)}, \quad k = 1,\ldots,p \\ & s_i^{-(k)}, s_r^{+(k)}, s_f^{o(k)}, s_g^{o(k)}, \lambda_j^{(k)} \geqslant 0,\ i \in I^{(k)},\ r \in O^{(k)},\ f \in M^{(k)}, \\ & \quad g \in N^{(k)}, \quad j = 1,\ldots,n, \quad k = 1,\ldots,p, \end{aligned} \tag{3}$$

where the non-Archimedean number ε in the objective function has been omitted for simplicity of expression. Note that if an input $X_c^{(k)}$ of process k is the same as one of its input intermediate products, $Z_c^{(k)}$, then they should be aggregated in the above formulation. This concept also applies to outputs (Färe & Grosskopf, 2000).

If an output efficiency is desired, then θ in the first constraint is removed and a variable ϕ is attached to $Y_{r0}^{(k)}$ in the second