DATA ENVELOPMENT ANALYSIS
Part I: Basic Theory and Mathematical Programming Models

Cesar G. Tapia
University of the Philippines Diliman
Quezon City

Abstract

Data envelopment analysis (DEA) is one of the most recently developed methods for performance evaluation of decision-making units (DMUs) in a given system where all DMUs utilize a common set of resources and produce a common set of outputs. This paper is the first in a series of papers on DEA particularly intended to reach Philippine readership. It introduces DEA's basic concepts and mathematical models used for efficiency evaluation.

1. Introduction

Performance evaluation in a given collection of decision-making units (DMUs) is the central concern of data envelopment analysis. DEA's main tool is mathematical programming, which is one of two known methods for efficiency analysis. It is a non-parametric method as opposed to parametric methods which are used in economics. The parametric methods rely on some explicit production function of outputs in terms of a combination of inputs.

Using a production function is helpful in explaining the input-output relationship. However, there is some difficulty in obtaining an explicit formulation for a production function because it can at best be assumed only a priori and can only be constructed from a set of empirical data (Seiford and Thrall (1990)).

Data envelopment analysis was initiated by Charnes, Cooper and Rhodes (1978). DEA is particularly useful for a system of DMUs which all have a common set of (one or more) input attributes together with a common set of (one or more) output attributes. The attributes may not necessarily have the same units of measure but are always assumed to have non-negative empirical data levels (or values). Referring to such a system of DMUs and utilizing a generalization of the classical definition of the engineering concept of efficiency as a ratio of a single output over a single input, the first mathematical formulation popularly known as the CCR ratio model involves a reduction of a multiple-output over a multiple-input situation to that of a single "virtual" output.
(combination) over a single "virtual" input. This is made possible by forming a linear combination of the multiple outputs and a linear combination of the multiple inputs for each of the DMUs, and then considering the value of the following ratio of the linear combination of outputs over that of the inputs as a measure of efficiency of the k-th DMU:

$$\sum_{j} v_j y_j^f / \sum_{i} u_i x_i^k$$

where $x_i^k$ ($i = 1, \ldots, n$) are known empirical input values of $k$ different input attributes of the k-th DMU; $y_j^k$ ($j = 1, \ldots, m$) are known empirical output values of $j$ different output attributes of the k-th ($k \in \{1, \ldots, K\}$) DMU; and $u_i$ and $v_j$ are variable linear combination coefficients for the input and output values, respectively. Charnes, Cooper and Rhodes intended to use this ratio as an efficiency measure for the k-th DMU. They required this ratio to have a value not to exceed unity. With this requirement, a DMU for which this ratio equals unity shall have an efficiency rating of one and this DMU is said to be efficient; otherwise, if the value of this ratio is less than unity, then the DMU is said to be inefficient. Charnes, Cooper and Rhodes also considered the necessity of imposing this requirement on all the DMUs. In other words, they required a set of normalizing conditions to be applied to all the DMUs' input and output values, which may be of different units of measures.

Thus, the CCR ratio model for a given DMU $k$ is given symbolically as:

$$(\text{CCR ratio model}) \quad \text{Max} \sum_{j} v_j y_j^f / \sum_{i} u_i x_i^k \quad \text{s.t.} \sum_{j} v_j y_j^f / \sum_{i} u_i x_i^k \leq 1, \forall t = 1, 2, \ldots, K \quad u_i, v_j \geq 0, \forall i, j$$

By means of the non-negative variable linear combination coefficients $u_i$ and $v_j$ for the multiple inputs and outputs respectively, the CCR model intends to determine the maximum value of this ratio for the k-th DMU subject to the satisfaction of all the normalizing requirements (one for each DMU). The optimal values of these coefficients $u_i^*$ and $v_j^*$ shall identify the maximum value of this ratio which is then interpreted as the efficiency rating of the k-th DMU.

Not requiring a'priori weights in evaluating the efficiency of a DMU, the CCR model was built on the earlier work of Farrell (1957) who developed a technical efficiency measure by means of a linear production function defined under the assumptions of constant returns-to-scale of a single-input-single-output system.

2 MATIMYAS MATHEMATIKA JAN 1984
The importance and applicability of data envelopment concepts and techniques are manifested by the sustained interest of researchers in conducting theoretical and empirical DEA studies. The empirical orientation of DEA and the absence of a priori assumptions have provided convenient applications in a number of studies involving efficiency frontier estimation in both the public and private sectors mostly in industrialized countries.

There are well over 400 articles on DEA published in international journals of management science, economics, mathematics and operations research. A comprehensive bibliography is given in Seiford (1990).

Recognizing our problem in the Philippines of very limited access to many popular journals, this author therefore writes this paper as the first in a series of papers. This will provide Philippine readership with the opportunity to learn not only about DEA's concepts and applications but also some recent research directions being pursued by other authors in general and by this author in particular.

Section 2 describes some basic DEA concepts. To illustrate these concepts, Section 3 describes a hypothetical system of five DMUs where the analysis for each DMU utilizes two input attributes and only one output attribute. This simple illustration allows for some geometric portrayal of the basic DEA concepts, particularly data envelopment within the efficiency frontier. Section 4 states some DEA applications and research directions done by other authors. Section 5 contains some concluding remarks.

2. Basic Concepts and Mathematical Programming Models

The CCR ratio model yields an infinite number of optimal solutions since if \((u_i, v_j)\) is optimal then \((\alpha u_i, \alpha v_j)\) is also optimal for \(\alpha > 0\). The transformation developed by Charnes and Cooper (1962) for linear fractional programming selects a representative solution for which \(\sum u_i x_i^k = 1\). Thus, the following model represents the equivalent linear programming problem which is the result of the transformation.

\[
\begin{align*}
\text{(MF)} \quad & \text{Max } u^T Y^k \\
\text{s.t.} \quad & v^T Y - u^T X \leq \Omega \\
& u^T X^k = 1 \\
& u_i, v_j \geq \epsilon, \quad \forall i, j
\end{align*}
\]

where \(X\) is a matrix consisting of \(K\) input vectors \(X^k (k = 1, \ldots, K)\), and \(Y\) is a matrix consisting of \(K\) output vectors \(Y^k (k = 1, \ldots, K)\); \(\epsilon\) is an infinitesimal number having a
non-Archimedean value in the order of $10^{-6}$ to $10^{-8}$.

The notations $u^T$ and $v^T$ denote the transpose of the column variable vectors $u$ and $v$ respectively, with components $u_i (i = 1, \ldots, I)$ and $v_j (j = 1, \ldots, J)$.

The dual formulation of this linear model for the $k$-th DMU is as follows:

\[
\text{(DF) } \quad \begin{align*}
\min \ z^k &= \theta_k \\
\text{s.t. } \theta_k X^k &\geq X \lambda \\
Y \lambda &\geq Y^k \\
\lambda &\geq 0
\end{align*}
\]

where the variables are the scalar $\theta_k$ and the $K$-column vector $\lambda$. $0$ is a $K$-column zero vector.

The dual formulation directly provides the efficiency rating of the $k$-th DMU as the optimal value of $\theta_k$ which is also known as the input reduction coefficient.

A DMU is defined to be efficient if and only if from the solution of the model DF, the optimal value of $\theta_k = 1$ (i.e., no reduction is necessary) and the slacks are all equal to zero for both the input and output attributes. Therefore a DMU $k$ is inefficient if $\theta_k < 1$. It is on the efficiency frontier if $\theta_k = 1$ and the input slacks and outputs slacks are all equal to zero.

Also from the dual formulation DF, the optimal value of the vector $\lambda$ identifies a subset called a local efficiency facet or simply a facet of efficient DMUs defining a boundary (otherwise known as the enveloping surfaces) for the $k$-th DMU's input and output empirical levels. In other words, the efficiency levels for the $k$-th DMU are specified by the product $X \lambda'^o$ for its inputs and by $Y \lambda'^o$ for its outputs. In particular, these efficiency levels are the lower bounds and the upper bounds for its inputs and outputs, respectively. The situation described above suitably demonstrates the idea of data envelopment.

To implement the dual formulation (DF), Charnes, Cooper and Rhodes (popularly known as CCR) proposed the following input-oriented model:

\[
\text{(CCR)} \quad \begin{align*}
\min \ z^k &= \theta_k - \epsilon \left[ \sum_i ax_i + \sum_j ay_j \right]
\end{align*}
\]
\[ s.t. \quad \theta_i x_i^b - s x_i = X_i \lambda, \quad \forall i = 1, \ldots, I \]
\[ \lambda, \varphi_j = 0 \]
\[ x_i, s_i, y_j \geq 0, \quad \forall i, j \]

where \( s_i \) and \( y_j \) are input and output slacks respectively, \( \varepsilon \) is an infinitesimal number of non-Archimedean value which pre-empts the minimization on the slacks before the minimization on the reduction factor \( \delta_k \); \( X_i \) is the \( i \)-th row of the input matrix \( X \); and \( \varphi_j \) is the \( j \)-th row of the output matrix \( Y \).

However, some difficulty can be experienced in the choice of value for \( \varepsilon \), and Ali and Seiford (1991) report computational inaccuracies and erroneous results from the use of certain \( \varepsilon \) non-Archimedean values for \( \varepsilon \). To avoid this problem they suggested a two-stage process of minimization where \( \delta_k \) is minimized first in the first stage, and the optimal value of \( \delta_k \) thus obtained is used in the second stage to identify the corresponding input and output slacks.

There are two basic types of envelopment surfaces in DEA: the constant returns-to-scale (CRS) and the variable-returns-to-scale (VRS). In the CRS, an increase in input requires a proportional increase in output; while in the VRS, the relative increase may not be proportional.

The model DF produces a CRS envelopment surface. If an additional constraint called the convexity requirement:

\[ \lambda \mathbf{1} = 1, \]

(where \( \lambda \) is a \( K \)-row vector with all components equal to unity) is included in the model DF, then the resulting DEA model produces a VRS envelopment surface. Both CRS and VRS envelopment surfaces are piecewise linear surfaces. Analysis of envelopment surfaces has led to significant extensions of other DEA theories, applications and interpretations (Charnes, et. al. 1982, 1983).

For purposes of identifying the efficiency levels of each of the DMUs under study, each of the basic DEA models can be formulated on the basis of the orientation of analysis which can be either input-oriented or output-oriented. For the input-oriented analysis, the focus is on the identification of possible reduction of the input values; while
for the output-oriented analysis, the focus is on the identification of possible augmentation of the output values.

From the dual DEA models such as the CCR model described above, the local efficiency surface which defines some lower bounds for the inputs of a DMU and the local efficiency surface which defines some upper bounds for its outputs are given respectively by $\mathbf{X}^{\text{lo}}$ and $\mathbf{X}^{\text{hi}}$. From the primal DEA models such as the MF model described above, they are given respectively by $(u^{\text{lo}})^T\mathbf{X}$ and $(v^{\text{hi}})^T\mathbf{Y}$. The local efficiency surfaces in both the input and output spaces are defined by a subset called a facet of efficient DMUs. From the dual models in particular, a facet is a subset \{DMU / $\lambda^{\text{lo}}_{\mathbf{X}} > 0$\} of efficient DMUs which correspond to the positive components of the vector $\lambda^{\text{lo}}_{\mathbf{X}}$.

For the evaluation of each DMU, the local efficiency facet for one DMU may be different from that for another DMU. This means that in using a DEA model such as CCR or DF, a DMU's efficiency rating is obtained by comparing each of its input and output levels with those of the linear combination of the corresponding levels of input and output of the efficient DMUs belonging to its local efficiency facet. Consequently, the efficiency ratings for the DMUs do not have a global or system significance since it is possible that the efficiency ratings may be based on two or more different local efficiency facets. This means that there is more than one efficiency standard identified in the system of DMUs. Ideally however, for a systematic (global) efficiency measurement of all the DMUs, there should be one efficiency standard applied to all DMUs.

3. A Hypothetical Example

Let us consider five DMUs having two input attributes $X_1$ and $X_2$ and only one output attribute $Y_1$. For the sake of simplicity as well as for the possibility of geometric portrayal, let us assume that the five DMUs all have the same one unit of output, while they have different values for the two input attributes. This assumption will allow us to look at the 2-Dimensional Input Space for our system of five DMUs without having to bother ourselves with the 1-Dimensional Output Space for the same system of DMUs. Thus the DMUs can be represented as in the figure below as points with coordinates $(x_1^k, x_2^k)$ for $k = 1, 2, \ldots, 5$.

The figure shows that DMU1, DMU2, DMU3, and DMU4 are situated on some efficiency frontier (lower boundary) for the various input levels in the system. For example, for DMU1 there is no linear combination (per input attribute) of the input levels
of other DMUs that is better (less in all attributes) then the corresponding input levels of DMU1. This is also shown from the result of the optimization of an input-oriented model such as the model DF described above. Such optimization yields Min \( z_1 = 6_1 = 1 \), which means that there is no reduction necessary for the input levels (0.75,5) of DMU1. This no reduction situation holds similarly for DMU2, DMU3, and DMU4. Under these situations, the optimal linear combination vectors \( x_{\text{OPT}}^k \) for DMU1 \( (k = 1) \), for DMU2 \( (k = 2) \), for DMU3 \( (k = 3) \), and for DMU4 \( (k = 4) \) are respectively: \( (1,0,0,0,0)^T \), \( (0,1,0,0,0)^T \), \( (0,0,1,0,0)^T \), and \( (0,0,0,1,0)^T \). These vectors indicate that such efficient DMUs are self-evaluating. The efficiency frontier is composed of the piecewise-linear facets formed by the efficient DMUs, DMU1, DMU2, DMU3, and DMU4.

Let us now consider the case of DMU5. The figure shows that there is a virtual DMU \( (3.59,1.47) \) on the efficiency facet defined in terms of all possible linear combinations of DMU3's and DMU4's input levels. This suggests that DMU5 is inefficient since it could actually use, instead of \( (4.25,1.75) \), the input levels \( (3.59,1.47) \) of the virtual DMU on this efficiency facet. This is also shown from the solution of the model DF, as follows:

\[
\text{Min } z^5 = 6^5 = 0.84211
\]
\[
\begin{align*}
\lambda_{\text{min}} &= \begin{bmatrix} 0.75 & 1 & 2 & 5 & 4.25 \\ 5 & 2 & 2 & 1 & 1.75 \end{bmatrix} \begin{bmatrix} 0 \\ 0.47 \\ 0.23 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.59 \\ 1.47 \end{bmatrix} = \gamma_{\text{max}} = 0.84211 \begin{bmatrix} 4.25 \\ 1.75 \end{bmatrix} \\
\gamma_{\text{min}} &= \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0.47 \\ 0.23 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} = \gamma^5 = \begin{bmatrix} 1 \end{bmatrix}_{1 \times 1}
\end{align*}
\]

The efficiency rating of DMU5 obtained from the solution of the model DF is actually \( \min \ z^2 = \delta_{\text{max}} = 0.84211 < 1 \). This value indicates the reduction coefficient which is necessary to bring simultaneously the current input levels \((4.25, 1.75)\) for DMU5 to the optimum (efficiency) levels \((3.59, 1.47)\). The optimal linear combination vector \( \lambda_{\text{min}} \) = \( (0, 0, 0, 0, 0.53, 0.53) \). The non-zero components of this optimal vector indicate the corresponding DMUs (i.e., DMU3 and DMU4) whose input levels are used as a basis for the efficiency rating for DMU5. On the other hand, the components of this vector corresponding to DMU1, DMU2, and DMU5 are zeroes. This means that their input levels are not significant components of the linear combinations \( \lambda_{\text{min}} \) and \( \gamma_{\text{min}} \) which determine the minimum value of the reduction coefficient \( \delta_{\text{max}} \) for DMU5.

The above illustration accompanied by some geometric portrayal can help in the understanding of the basic concepts of DEA. It should be noted, however, that any number of attributes beyond two or three shall mean a corresponding number of dimensions in either the input or output spaces. This can make the geometric representation very difficult if not impossible.

4. DEA Applications

Numerous DEA applications in the public sector, the regulated sector and the private sector are described in the literature.

Among the published applications in the public sector are the following. (1)
Technical efficiency evaluation of New Jersey hospitals (Borden, 1988); (2) US public school education evaluation (Charnes, Cooper and Rhodes, 1981; Bessent et. al., 1983); (3) evaluation of colleges and universities (Ahn, Charnes and Cooper, 1988a; Rhodes and Southwick, 1989); (4) railroad property evaluation (Adolphson, Cornia and Walters, 1989); (5) evaluation of maintenance units of the US Air Force (Charnes, et. al., 1985); (6) analysis of management of army recruiting districts (Ali and Seiford, 1991); (7) highway patrol evaluation (Cook et. al., 1991); and (8) economic performance evaluation of Chinese cities (li, 1989; Sueyoshi, 1991).

In the regulated sector, published DEA applications are (1) evaluation of electric cooperatives (Thomas, 1985); and (2) evaluation of some not-for-profit organizations (Ahn et. al., 1988b); among others.

In the private sector, two among several applications are the evaluation of audits of (1) banks (Sun, 1988; Charnes et. al., 1990) and colleges and universities (Ahn, Charnes and Cooper, 1988a).

In all of the above applications as well as in future applications, DEA is useful in establishing certain efficiency criteria on the basis of which the performance of each DMU can be measured. Thus, the efficient DMUs as well as the inefficient ones can be identified. The management of the DMUs can therefore have a guideline resulting from DEA. Such guidelines, however, has only a local significance - in the sense that it is drawn from a local efficiency facet. Nevertheless, this is useful information for the individual managers of the DMUs since they may adopt policies differently according to each one of their own specific requirements, and perhaps such policies may be dictated only by the performance standard set locally among some "neighboring" DMUs.

In the second paper, we shall address the need for establishing a system standard for evaluating DMUs. This will particularly be useful for a system of DMUs which are all controlled and regulated by a central decision-making body. As far as a system-wise evaluation is concerned, the efficiency ratings produced by existing DEA methodologies should not be globally compared unless they are the results of evaluation based on a single efficiency facet. This is because different local efficiency facets actually represent different efficiency standards.

Therefore, in the next paper this author will propose an algorithm that may be used for constructing a unique efficiency facet for all the DMUs in the system under study. This will be necessary when centralized (global) efficiency evaluation is desired in lieu of a local evaluation.

Researchers have worked on some extensions of the basic DEA concepts such as the
following: (1) the hierarchical organizations of DMUs where "smaller" DMUs can be evaluated against those identified with a set of "larger" category DMUs; (2) the non-discretionary variables' role in DEA as variables which cannot be controlled by the decision-maker; (3) the global non-comparability of DMUs in cases where the efficiency is composed of two or more local efficiency facets; and (4) the multiple reduction for the various inputs in lieu of the proportional reduction for all inputs in the basic DEA models.

4. Conclusion

In this paper, which is the first in a series of papers designed for Philippine readership, the reader is introduced to some basic concepts and mathematical programming models that for just over fifteen (15) years have been widely used in developing a non-parametric method known as data envelopment analysis (DEA) for performance evaluation of decision-making units (DMUs). Actual DEA applications are enumerated to give the reader an idea of the wide acceptance of DEA specifically in the USA. Through this paper, the reader is made aware of the fact that the results of existing DEA generally have little system-wise significance because of the model's inherent characteristics. The reader is therefore foretold of the intention to address in the next paper the issue of global non-comparability of existing DEA efficiency ratings.

References

Ahn, T., A. Charnes and W.W. Cooper, 1988a. "Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning", Socio-

Ahn, T., A. Charnes and W.W. Cooper, 1988b. "Using DEA to measure the efficiency of not-for-profit organizations: a critical evaluation-comment", Managerial and


Science, Vol 30 No 9, pp 1078-1092.


Thomas, D.I., 1983. "Auditing the efficiency of regulated companies through the use of DEA: an application to electric cooperatives". Ph.D. Thesis, The University of Texas at Austin, Graduate School of Business (Available from University Microfilms, Inc., 300 North Zeeb Road, Ann Arbor, MI 48106).