European Journal of Operational Research 142 (2002) 16–20

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Continuous Optimization

Modeling undesirable factors in efficiency evaluation

Lawrence M. Seiford [a], Joe Zhu [b,*]

[a] Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117, USA
[b] Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Received 3 October 2000; accepted 9 August 2001

Abstract

Data envelopment analysis (DEA) measures the relative efficiency of decision making units (DMUs) with multiple performance factors which are grouped into outputs and inputs. Once the efficient frontier is determined, inefficient DMUs can improve their performance to reach the efficient frontier by either increasing their current output levels or decreasing their current input levels. However, both desirable (good) and undesirable (bad) factors may be present. For example, if inefficiency exists in production processes where final products are manufactured with a production of wastes and pollutants, the outputs of wastes and pollutants are undesirable and should be reduced to improve the performance. Using the classification invariance property, we show that the standard DEA model can be used to improve the performance via increasing the desirable outputs and decreasing the undesirable outputs. The method can also be applied to situations when some inputs need to be increased to improve the performance. The linearity and convexity of DEA are preserved through our proposal.

Keywords: Data envelopment analysis (DEA); Classification invariance; Efficiency; Undesirable factors

1. Introduction

Data envelopment analysis (DEA) uses linear programming problems to evaluate the relative efficiencies and inefficiencies of peer decision making units (DMUs) which produce multiple outputs and multiple inputs. Once DEA identifies the efficient frontier, DEA improves the performance of inefficient DMUs by either increasing the current output levels or decreasing the current input levels. However, both desirable (good) and undesirable (bad) output and input factors may be present. Consider a paper mill production where paper is produced with undesirable outputs of pollutants such as biochemical oxygen demand, suspended solids, particulates and sulfur oxides. If inefficiency exists in the production, the undesirable pollutants should be reduced to improve the inefficiency, i.e., the undesirable and desirable outputs should be treated differently when we evaluate the production performance of paper mills. However, in the standard DEA model, decreases in outputs are not allowed and only inputs are allowed to decrease. (Similarly, increases in

* Corresponding author. Tel.: +1-508-831-5467; fax: +1-508-831-5720.
E-mail addresses: seiford@umich.edu (L.M. Seiford), jzhu@wpi.edu (J. Zhu).

0377-2217/02/$ - see front matter
PII: S0377-2217(01)00293-4

inputs are not allowed and only outputs are allowed to increase.) If one treats the undesirable outputs as inputs, the resulting DEA model does not reflect the true production process. Färe et al. (1989) develop a non-linear DEA program to model the paper production system where the desirable outputs are increased and the undesirable outputs are decreased.

Situations when some inputs need to be increased to improve the performance are also likely to occur. For example, in order to improve the performance of a waste treatment process, the amount of waste (undesirable input) to be treated should be increased rather than decreased as assumed in the standard DEA model.

The current paper develops an alternative approach to treat both desirable and undesirable factors differently in the standard linear BCC DEA model of Banker et al. (1984). This preserves the linearity and convexity in the BCC model. The key to our approach is the use of DEA classification invariance under which classifications of efficiencies and inefficiencies are invariant to the data transformation.

2. Background

A DEA data domain can be characterized by a data matrix

$$P = \begin{bmatrix} Y \\ -X \end{bmatrix} = [P_1, \ldots, P_n]$$

with $s+m$ rows and n columns. Each column corresponds to one of the DMUs. The jth column

$$P_j = \begin{bmatrix} Y_j \\ -X_j \end{bmatrix}$$

is composed of an input vector x_j whose ith component x_{ij} is the amount of input i used by DMU_j and an output vector y_j whose rth component y_{rj} is the amount of output r produced by DMU_j.

The BCC efficiency can be obtained by calculating the following linear programming problem:

$$\begin{aligned} \max \quad & \eta \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_j x_j + s^- = x_o, \\ & \sum_{j=1}^{n} z_j y_j - s^+ = \eta y_o, \\ & \sum_{j=1}^{n} z_j = 1, \\ & z_j \geqslant 0, \quad j = 1, \ldots, n, \end{aligned} \tag{1}$$

where x_o and y_o represent the input and output vectors of DMU_o under evaluation. This model is an output-oriented program. Similarly, one can write an input-oriented BCC model

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_j x_j + s^- = \theta x_o, \\ & \sum_{j=1}^{n} z_j y_j - s^+ = y_o, \\ & \sum_{j=1}^{n} z_j = 1, \\ & z_j \geqslant 0, \quad j = 1, \ldots, n. \end{aligned} \tag{2}$$

Next suppose the input vector is displaced by the m rowed vector u and the output vector is displaced by the s rowed vector v. That is, $\bar{x}_j = x_j + u$ and $\bar{y}_j = y_j + v$ $(j = 1, 2, \ldots, n)$.

Ali and Seiford (1990) provide the following result concerning the translation invariance in the BCC model:

Classification invariance. DMU_o is efficient under (1) or (2) if and only if DMU_o is efficient under (1) or (2) with translated data; DMU_o is inefficient under (1) or (2) if and only if DMU_o is inefficient under (1) or (2) with translated data.

In general, there are three cases of invariance under data transformation in DEA. The first case is restricted to the "classification invariance" where the classifications of efficiencies and inefficiencies are invariant to the data transformation. The second case is the "ordering invariance" of the inefficient DMUs. The last case is the "solution invariance" in which the new DEA model (after data translation) must be equivalent to the old one, i.e., both mathematical programming problems must have exactly the same solution. The current paper is concerned only with the first level of invariance – classification invariance. See Pastor (1996) and Lovell and Pastor (1995) for recent developments in invariance property in DEA.

3. Undesirable factors in DEA

Suppose the DEA data domain is expressed as

$$\begin{bmatrix} Y \\ -X \end{bmatrix} = \begin{bmatrix} Y^{\mathrm{g}} \\ Y^{\mathrm{b}} \\ -X \end{bmatrix}, \tag{3}$$

where Y^{g} and Y^{b} represent the desirable (good) and undesirable (bad) outputs, respectively.

Obviously, we wish to increase the Y^{g} and to decrease the Y^{b} to improve the performance. However, in the standard BCC model (1), both Y^{g} and Y^{b} are supposed to increase to improve the performance. In order to increase the desirable outputs and to decrease the undesirable outputs, Färe et al. (1989) modify the model (1) into the following non-linear programming problem:

$$\begin{aligned} \max \quad & \Gamma \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_j x_j + s^- = x_{\mathrm{o}}, \\ & \sum_{j=1}^{n} z_j y_j^{\mathrm{g}} - s^+ = \Gamma y_{\mathrm{o}}^{\mathrm{g}}, \\ & \sum_{j=1}^{n} z_j y_j^{\mathrm{b}} - s^+ = \frac{1}{\Gamma} y_{\mathrm{o}}^{\mathrm{b}}, \\ & \sum_{j=1}^{n} z_j = 1, \\ & z_j \geqslant 0, \quad j = 1, \ldots, n. \end{aligned} \tag{4}$$

Based upon classification invariance, we next show that an alternative to model (4) can be developed to preserve the linearity and convexity in DEA.

First we multiply each undesirable output by "−1" and then find a proper translation vector w to let all negative undesirable outputs be positive. The data domain of (3) now becomes

$$\begin{bmatrix} Y \\ -X \end{bmatrix} = \begin{bmatrix} Y^{\mathrm{g}} \\ \bar{Y}^{\mathrm{b}} \\ -X \end{bmatrix}, \tag{5}$$

where the jth column of (translated) bad output now is $\bar{y}_j^{\mathrm{b}} = -y_j^{\mathrm{b}} + w > 0$.

Based upon (5), model (1) becomes the following linear program:

$$\begin{aligned} \max \quad & h \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_j y_j^{\mathrm{g}} \geqslant h y_{\mathrm{o}}^{\mathrm{g}}, \\ & \sum_{j=1}^{n} z_j \bar{y}_j^{\mathrm{b}} \geqslant h \bar{y}_{\mathrm{o}}^{\mathrm{b}}, \\ & \sum_{j=1}^{n} z_j x_j \leqslant x_{\mathrm{o}}, \\ & \sum_{j=1}^{n} z_j = 1, \\ & z_j \geqslant 0, \quad j = 1, \ldots, n. \end{aligned} \tag{6}$$

Note that (6) expands desirable outputs and contracts undesirable outputs as in the non-linear DEA model (4).

The following theorem ensures that the optimized undesirable output of $y_{\mathrm{o}}^{\mathrm{b}}$ $(= w - h^* \bar{y}_{\mathrm{o}}^{\mathrm{b}})$ cannot be negative:

Theorem 1. *Given a translation vector* w, *suppose* h^* *is the optimal value to* (6), *we have*

$h^* \bar{y}_{\mathrm{o}}^{\mathrm{b}} \leqslant w.$

Proof. Note that all outputs now are non-negative. Let z_j^* be an optimal solution associated with h^*. Since $\sum_{j=1}^{n} z_j^* = 1$, therefore $h^* \bar{y}_0^{\mathrm{b}} \leqslant \bar{y}^*$, where $\bar{y}^*$ is composed from (translated) maximum values among all bad outputs. Note that $\bar{y}^* = -\bar{y}^* + w$, where $\bar{y}^*$ is composed from (original) minimum values among all bad outputs. Thus, $h^* \bar{y}_{\mathrm{o}}^{\mathrm{b}} \leqslant w$. □

There are indeed at least five possibilities for dealing with undesirable outputs in the DEA-BCC framework. The first possibility is just simply to ignore the undesirable outputs. The second is to treat the undesirable outputs in the non-linear DEA model (4). The third is to treat the undesir-

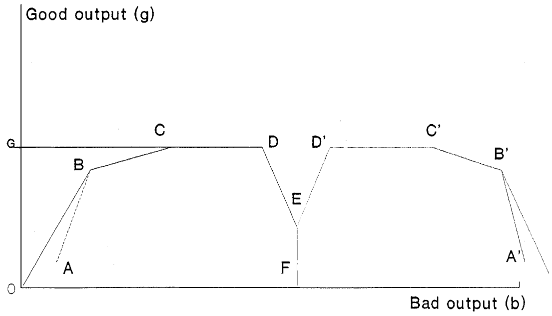

Fig. 1. Treatment of bad output.

able ones as outputs and to adjust the distance measurement in order to restrict the expansion of the undesirable outputs (see the weak disposability model in Färe et al., 1989). The fourth is to treat the undesirable outputs as inputs. However, this does not reflect the true production process. The fifth is to apply a monotone decreasing transformation (e.g., $1/y^{\mathrm{b}}$) to the undesirable outputs and then to use the adapted variables as outputs. The current paper, in fact, applies a linear monotone decreasing transformation. Since the use of linear transformation preserves the convexity relations, it is a good choice for a DEA model. [1]

Fig. 1 illustrates the last three methods for treating the undesirable outputs. The five DMUs A, B, C, D and E use an equal input to produce one desirable output (g) and one undesirable output (b). The region OGCDEF is the conventional output set under the output-oriented BCC model (1). If we suppose weak disposability of undesirable output, then the output set is the region OBCDEF in which feasible radial contractions rather than feasible vertical extensions are allowed to the origin. If we treat the undesirable output as an input, then ABCD becomes the BCC frontier. For the fifth method, by a proper translation vector, we may rotate the output set at EF and obtain the symmetrical region. In this case, DMUs A′, B′ and C′, which are, respectively, the adapted points of A, B and C, are efficient.

The above discussion can also be applied to situations when some inputs need to be increased rather than decreased to improve the performance. In this case, we rewrite the DEA data domain as

$$\begin{bmatrix} Y \\ -X \end{bmatrix} = \begin{bmatrix} Y \\ -X^{\mathrm{I}} \\ -X^{\mathrm{D}} \end{bmatrix}, \tag{7}$$

where X^{I} and X^{D} represent inputs to be increased and decreased, respectively.

Next multiply X^{I} by "−1" and then find a proper translation vector k to let all negative X^{I} be positive. The data domain of (7) becomes

$$\begin{bmatrix} Y \\ -X \end{bmatrix} = \begin{bmatrix} Y \\ -\bar{X}^{\mathrm{I}} \\ -X^{\mathrm{D}} \end{bmatrix}, \tag{8}$$

where the jth column of (translated) input to be increased now is $\bar{x}_j^{\mathrm{I}} = -x_j^{\mathrm{I}} + k > 0$.

Based upon (8) and model (2), we have

$$\begin{aligned} \min \quad & \tau \\ \text{s.t.} \quad & \sum_{j=1}^{n} z_j x_j^{\mathrm{D}} + s^- = \tau x_{\mathrm{o}}^{\mathrm{D}}, \\ & \sum_{j=1}^{n} z_j \bar{x}_j^{\mathrm{I}} + s^- = \tau \bar{x}_{\mathrm{o}}^{\mathrm{I}}, \\ & \sum_{j=1}^{n} z_j y_j - s^+ = y_{\mathrm{o}}, \\ & \sum_{j=1}^{n} z_j = 1, \\ & z_j \geqslant 0, \quad j = 1, \ldots, n, \end{aligned} \tag{9}$$

where X^{I} is increased and X^{D} is decreased for a DMU to improve the performance.

We conclude this section by applying our method to the 30 paper mills (Färe et al., 1989) that used fiber, energy, capital and labor as inputs to produce paper, together with four undesirable outputs (pollutants): biochemical oxygen demand, total suspended solids, particulates and sulfur oxides.

Table 1 gives the efficiency scores. Column 2 shows the optimal value to the model (1) when all pollutants are not included, i.e., the paper produced is used as the only output (undesirable outputs are ignored). Column 3 reports the efficiency scores obtained from model (6) with a

[1] See Lewis and Sexton (1999) for an alternative approach in treating inputs.

Table 1
Efficiency scores for the 30 paper mills

Mill No.	Model (1)[a]	h^* in model (6)	Model (1)[b]
1	1.21079	1.00000	1.08155
2	1.29546	1.00000	1.00000
3	1.00000	1.00000	1.00000
4	1.17613	1.00000	1.01678
5	1.51135	1.00146	1.38655
6	1.00000	1.00000	1.00017
7	1.00000	1.00000	1.00000
8	1.37233	1.01553	1.21051
9	1.00000	1.00000	1.00000
10	1.00000	1.00000	1.00000
11	1.11581	1.03276	1.11581
12	1.14381	1.05248	1.14381
13	1.38071	1.00180	1.37900
14	1.20791	1.02237	1.20791
15	1.00000	1.00000	1.00061
16	1.00000	1.00000	1.00000
17	1.00000	1.00000	1.00000
18	1.02972	1.00000	1.00000
19	1.44505	1.02877	1.44505
20	1.53663	1.02367	1.53663
21	1.00000	1.00000	1.00000
22	1.35036	1.00349	1.23445
23	1.24970	1.02677	1.22360
24	1.27092	1.00000	1.00000
25	1.00000	1.00000	1.00000
26	1.16223	1.00000	1.02932
27	1.00000	1.00000	1.00000
28	1.00000	1.00000	1.00000
29	1.01650	1.00000	1.00000
30	1.11780	1.02783	1.08411
Mean	1.15311	1.00790	1.10320

[a] Only paper produced is used as the output. All the undesirable outputs (pollutants) are ignored.
[b] The bad outputs are treated as inputs in model (1).

translation vector of $w = (20\,000, 10\,000, 3000, 15\,000)$. The last column gives the results obtained from (1) where all four undesirable outputs are treated as inputs.

When we ignore pollutants, 12 mills were deemed as efficient. However, we have 11 inefficient mills under model (6), namely, mills 5, 8, 11, 12, 13, 14, 19, 20, 22, 23. The above results confirm the finding in Färe et al. (1989) that failure to credit mills for pollution reduction can severely distort the ranking of mill performance.

4. Conclusion

Under the context of the BCC model, the current paper provides an alternative method in dealing with desirable and undesirable factors in DEA. As a result, convexity and linearity are preserved. On the basis of BCC classification invariance, a linear monotone decreasing transformation is applied to treat the undesirable outputs so that the output-oriented BCC model allows the expansion of desirable outputs and the contraction of undesirable outputs. The new approach can also be applied to situations when some inputs need to be increased to improve the performance.

Acknowledgements

The authors appreciate the comments by Professor J.T. Pastor and an anonymous referee and thank Professor Suthathip Yaisawarng for providing the data on the paper mills. The authors would also like to thank Professor Jyrki Wallenius.

References

Ali, A.I., Seiford, L.M., 1990. Translation invariance in data envelopment analysis. Operations Research Letters 9, 403–405.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30, 1078–1092.

Färe, R., Grosskopf, S., Lovell, C.A.K., Pasurka, C., 1989. Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. The Review of Economics and Statistics 71, 90–98.

Lewis, H.F., Sexton, T.R., 1999. Data envelopment analysis with reverse inputs, paper presented at North America Productivity Workshop, Union College, Schenectady, NY, July 1999.

Lovell, C.A.K., Pastor, J.T., 1995. Units invariant and translation invariant DEA models. Operations Research Letters 18, 147–151.

Pastor, T., 1996. Translation invariance in DEA: A generalization. Annals of Operations Research 66, 93–102.