ELSEVIER

European Journal of Operational Research 89 (1996) 473-481

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis

Rajiv D. Banker [a], Hsihui Chang [b], William W. Cooper [c,*]

[a] Carlson School of Management, University of Minnesota, Minneapolis, MN 55455, USA
[b] College of Commerce, National Chengchi University, Taipei, Taiwan, ROC
[c] Graduate School of Business, University of Texas at Austin, Austin, TX 78712-1174, USA

Received 29 March 1994; revised 17 January 1995

Abstract

This paper discusses alternative methods for determining returns to scale in DEA. The methods for estimating returns to scale in DEA, as developed by Banker (1984), Banker, Charnes and Cooper (1984) and Banker and Thrall (1992), are proved to be conceptually equivalent to the two-stage methods of Färe, Grosskopf and Lovell (1985) when their assumptions apply. Here the emphasis is on the CCR model of DEA and very simple methods are introduced for determining returns to scale locally with this model by reference to Banker's concept of Most Productive Scale Size.

Keywords: Data envelopment analysis; Returns to scale; Most productive scale size

1. Introduction

Charnes, Cooper and Rhodes (1978, 1981) introduced the CCR model of Data Envelopment Analysis (DEA) to evaluate the relative efficiency of decision making units (DMUs). Banker, Charnes and Cooper (1984) subsequently introduced the BCC model which separates technical efficiency and scale efficiency. Later, Banker (1984) showed how the CCR formulation can be employed to estimate most productive scale size (MPSS) and returns to scale (RTS). More recently, Banker and Thrall (1992) showed that the BCC and CCR methods of returns to scale estimation in Banker (1984) and Banker, Charnes and Cooper (1984) are equivalent.

Färe, Grosskopf and Lovell (1985) provided an alternative method for the estimation of returns to scale using DEA. Several recent studies (e.g., Chang and Guh, 1991) have regarded their estimation of returns to scale as being conceptually different from Banker et al. (Banker (1984); Banker, Charnes and Cooper, 1984; Banker and Thrall, 1992) and, indeed, Chang and Guh, 1991 characterize the Banker et al. method as invalid. However, in this paper we show that the two methods are equivalent [1]. Finally, we introduce a way to implement CCR model results which dispenses with the need for assumptions like 'strong' and 'weak disposal', as used by Färe, Grosskopf

* Corresponding author.

[1] See also Färe and Grosskopf and Banker, Chang and Cooper (1995) for simulation results that also support what is said here.

SSDI 0377-2217(95)00044-5

and Lovell (1985), as well as the need for assuming that technical efficiency has already been achieved, as in Banker and Thrall (1992). Numerical examples illustrate these uses.

2. DEA models and estimation of returns to scale

We begin by describing the BCC and CCR methods for estimating returns to scale in DEA. The following formulation is referred to as the BCC model (Banker, Charnes and Cooper, 1984):

$$h_o^B = \min h \tag{1.0}$$

subject to

$$\sum_{j=1}^{n} \lambda_j X_j \le h X_o, \tag{1.1}$$

$$\sum_{j=1}^{n} \lambda_j Y_j \ge Y_o, \tag{1.2}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \tag{1.3}$$

$$h, \lambda_j \ge 0, \tag{1.4}$$

where X_j and Y_j are vectors of observed inputs and outputs, respectively, for each of n DMUs, and X_o and Y_o represent the one to be evaluated [2].

The following is referred to as the CCR model (Charnes, Cooper and Rhodes, 1978):

$$h_o^C = \min h \tag{2.0}$$

subject to

$$\sum_{j=1}^{n} \lambda_j X_j \le h X_o, \tag{2.1}$$

$$\sum_{j=1}^{n} \lambda_j Y_j \ge Y_o, \tag{2.2}$$

$$h, \lambda_j \ge 0. \tag{2.4}$$

In the above two problems the BCC efficiency measure h_o^B provides a measure of technical efficiency and the CCR efficiency measure h_o^C is aggregate technical *and* scale efficiency. Then, as in (Banker, Charnes and Cooper, 1984), the scale efficiency h_o^S, is defined by the ratio h_o^C/h_o^B. Evidently $h_o^S = h_o^C/h_o^B \le 1$, which follows from $h_o^C \le h_o^B$ because the program in (1) is more constrained than (2).

Before proceeding we should note that we depart from Banker, Charnes and Cooper (1984), and other parts of DEA, by not attending to non-zero slack as sources of inefficiency in the above models. This means that any non-zero slack represents an excess that has no value and can be disposed of without cost or trouble. Given the name 'free disposal' by Koopmans (1957, p. 14) [3], it is further refined to 'strong' and 'weak' disposability by Färe, Grosskopf and Lovell, who use this assumption in their treatment of returns to scale. We do not elaborate further on this topic, however, but simply follow the usage of Färe, Grosskopf and Lovell by assigning zero coefficients to the slack variables in the objectives of the above models, just as is done in ordinary linear programming (Färe, Grosskopf and Lovell, 1985, pp. 178 ff.; Byrnes, Färe and Grosskopf, 1984, p. 672).

Banker (1984) shows how we can determine the situation for returns to scale locally at the levels represented in (X_o, Y_o) by examining the value of $\sum_{j=1}^{n} \lambda_j^*$, obtained from the CCR model when that sum is uniquely determined. [4] Banker and Thrall (1992) subsequently extended this to situations when alternative optimal solutions exist for $\sum_{j=1}^{n} \lambda_j^*$ – where '*' means an optimal value.

Formally, the BT conditions are as follows, where the reference is to returns to scale locally,

[2] In this paper we restrict attention to input-oriented models. See Banker et al. (1989) for other types.

[3] See also Koopmans (1951).

[4] We note that the occurrence of alternative optimal solutions with different values of $\sum_{j=1}^{n} \lambda_j^*$ is not very likely in empirical applications because this requires some of the CCR efficient observatons to be linearly dependent. This can be seen from Fig. 1 where for multiple optimal solutions (X^B, Y^B) and (X^C, Y^C), we must have $k(X^B, Y^B) = (X^C, Y^C)$ for some $k > 0$ (Zhu and Shen, 1992, 1993; Ganley and Cubbin, 1992).

CCR Returns to Scale Theorem (Banker and Thrall, 1992).

If $\sum_{j=1}^{n}\lambda_j^* = 1$ *in* any *alternate optima, then constant returns to scale prevail.*

If $\sum_{j=1}^{n}\lambda_j^* > 1$ *for* all *alternate optima, then decreasing returns to scale prevail.*

If $\sum_{j=1}^{n}\lambda_j^* < 1$ *for* all *alternate optima, then increasing returns to scale prevail.*

An alternative approach to the estimation of returns to scale is provided by Banker, Charnes and Cooper (1984) in terms of the dual to the linear program in (1), which we formulate as

$$h_o^B = \max UY_o + u_o \tag{3.0}$$

subject to

$$UY_j - VX_j + u_o \leq 0 \quad \text{for } j = 1,\dots,n, \tag{3.1}$$

$$VX_o = 1, \tag{3.2}$$

$$U \geq 0,\ V \geq 0, \text{ and } u_o \text{ is unconstrained in sign,} \tag{3.4}$$

where U and V are vectors of dual variables associated with the vectors Y_j and X_j, respectively.

If we have a unique optimal solution to (3), then constant returns to scale correspond to $u_o^* = 0$, increasing returns to scale correspond to $u_o^* > 0$, and decreasing returns to scale correspond to $u_o^* < 0$ considered over *all* alternate optima and when we are on the Efficient Production Frontier. [5] Banker and Thrall (1992, pp. 79-82) generalize this to the case of multiple optimal solutions. This is done by introducing new variables u_o^+ and u_o^- which represent optimal solutions obtained by solving (3) with one more constraint, viz.,

$$UY_o + u_o = 1, \tag{3.5}$$

after replacing X_j with hX_j, and replacing the objective function in (3.0) by either

$$u_o^+ = \max u_o \tag{3.1a}$$

or

$$u_o^- = \max - u_o. \tag{3.1b}$$

Banker and Thrall (1992) also prove the following theorem:

BCC Returns To Scale Theorem (Banker and Thrall, 1992). *Increasing returns to scale prevail at the point* (X_o, Y_o) *on the efficient production frontier if and only if* $u_o^- \geq u_o^+ > 0$; *constant returns to scale prevail at* (X_o, Y_o) *if and only if* $u_o^- \geq 0 \geq u_o^+$; *and decreasing returns to scale prevail at* (X_o, Y_o) *if and only if* $0 > u_o^- \geq u_o^+$.

Banker and Thrall (1992) also show that the alternative methods incorporated in the above two theorems are equivalent. Specifically, they show in Corollary 2 (p. 81) that the sign of u_o^* in (3) is the same as the sign of $(1 - \sum_{j=1}^{n}\lambda_j^*)$ in (2) for *all* optimal solutions if the CCR efficiency estimate is not equal to one and in *some* optimal solutions if it is equal to one. This alternative definition, based on the sign of u_o^*, focuses on the supporting hyperplane to the production possibility set, and highlights the fact that the concept of returns to scale is unambiguous only at points *on* the efficient sections of the production frontier. Clearly, when (X_o, Y_o) is an interior (inefficient) point, the estimated returns to scale may depend on whether we move to the frontier by contracting inputs or by augmenting outputs. See the discussion in Banker, Bardhan and Cooper (1996).

Färe, Grosskopf and Lovell (1985, p. 180) suggest the following two-step method to estimate returns to scale. In the first step, the BCC and the CCR models in (1) and (2) are solved to determine the scale efficiency h_o^S, as the ratio of h_o^C to h_o^B as in Banker, Charnes and Cooper (1984). A value of $h_o^S = 1$ indicates constant returns to scale at (X_o, Y_o) and a value of $h_o^S < 1$ indicates nonconstant returns to scale at (X_o, Y_o). In the second step, when $h_o^S < 1$, the following linear program is solved to determine whether

[5] Banker and Thrall (1992) assume that radial technical efficiency has been attained but this is not necessary in the uses we make of the CCR model. Banker, Bardhan and Cooper (1996) show how to eliminate this assumption when using the BCC model and we eliminate it here for the CCR model.

the scale inefficiency is associated with increasing or decreasing returns to scale:

$$h_o^F = \min h \tag{4.0}$$

subject to

$$\sum_{j=1}^{n} \lambda_j X_j \le h X_o, \tag{4.1}$$

$$\sum_{j=1}^{n} \lambda_j Y_j \ge Y_o, \tag{4.2}$$

$$\sum_{j=1}^{n} \lambda_j \le 1, \tag{4.3}$$

$$h, \lambda_j \ge 0. \tag{4.4}$$

We denote the DEA efficiency estimate obtained from (4) by h_o^F. Clearly, $h_o^C \le h_o^F \le h_o^B$ since the linear program in (4) is more constrained than that in (2) but less constrained than that in (1). Färe, Grosskopf and Lovell (1985) state that increasing returns to scale prevail if $h_o^S < 1$ and $h_o^F = h_o^B$, and decreasing returns to scale prevail if $h_o^S < 1$ and $h_o^F < h_o^B$.

Chang and Guh (1991), among others, appear to believe that this two-step FGL method is conceptually different from the Banker et al. method. Via the following theorem, however, we show that these alternative approaches for estimating returns to scale for any DMU are equivalent.

Theorem. (a) *There exists a solution of the CCR model with* $\sum_{j=1}^{n} \lambda_j^* = 1$ *if and only if* $h_o^S = 1$. (b) *All alternative optimal solutions of the CCR model have* $\sum_{j=1}^{n} \lambda_j^* > 1$ *if and only if* $h_o^S < 1$ *and* $h_o^F = h_o^B$. (c) *All optimal solutions of the CCR model have* $\sum_{j=1}^{n} \lambda_j^* < 1$ *if and only if* $h_o^S < 1$ *and* $h_o^F < h_o^B$.

Proof of Theorem. (a) If there exists a solution to the CCR model in (2) with $\sum_{j=1}^{n} \lambda_j^* = 1$, the CCR solution is feasible for the BCC model in (1) because the additional constraint $\sum_{j=1}^{n} \lambda_j = 1$ in the BCC model is obviously satisfied. Therefore, $h_o^C = h_o^B$ and $h_o^S = 1$. Conversely, if $h_o^S = 1$ and $h_o^C = h_o^B$, then a BCC solution also solves the CCR problem with $\sum_{j=1}^{n} \lambda_j^* = 1$. Therefore, there exists an optimal solution with $\sum_{j=1}^{n} \lambda_j^* = 1$ if and only if $h_o^S = 1$. This proves part (a) of the theorem.

We require the following lemma to prove the remaining two parts of the theorem:

Lemma. *If there exist two alternative optimal solutions to the CCR program in* (2), *one with* $\sum_{j=1}^{n} \lambda_j^{*(1)} < 1$ *and the other with* $\sum_{j=1}^{n} \lambda_j^{*(2)} > 1$, *then there exists an alternative optimal solution to* (2) *with* $\sum_{j=1}^{n} \lambda_j^{*(3)} = 1$.

Proof of Lemma. Let $\alpha_1 = \sum_{j=1}^{n} \lambda_j^{*(1)} > 1$ and $\alpha_2 = \sum_{j=1}^{n} \lambda_j^{*(2)} < 1$. Define $\alpha_3 = (\alpha_1 - 1)/(\alpha_1 - \alpha_2)$ so that $0 < \alpha_3 < 1$. Next, define

$$\lambda_j^{*(3)} = (1 - \alpha_3)\lambda_j^{*(1)} + \alpha_3 \lambda_j^{*(2)}, \quad j = 1, \dots, n.$$

Then it is easy to verify that $\{h_o^C, \lambda_1^{*(3)}, \dots, \lambda_j^{*(3)}, \dots, \lambda_n^{*(3)}\}$ constitutes an optimal solution to the CCR program in (2), and

$$\sum_{j=1}^{n} \lambda_j^{*(3)} = (1 - \alpha_3) \sum_{j=1}^{n} \lambda_j^{*(1)} + \alpha_3 \sum_{j=1}^{n} \lambda_j^{*(2)} = 1. \quad \square$$

Proof of Theorem (continued). With this lemma in hand we now move to proof of parts (b) and (c) in the above theorem as follows: We start with part (c). If $\sum_{j=1}^{n} \lambda^* < 1$ for all alternative solutions to the CCR program in (2), then it follows from part (a) of the theorem that $h_o^C < h_o^B$. Also, in this case, the solutions to the CCR program are feasible for the linear program in (4) and therefore, $h_o^F = h_o^C < h_o^B$. It also follows from the lemma that if $h_o^S < 1$ and there exists no optimal solution with $\sum_{j=1}^{n} \lambda_j^* \le 1$, then we must have $\sum_{j=1}^{n} \lambda_j^* > 1$ for all alternative optimal solutions. This proves part (c) of the theorem.

Next we turn to part (b) of the theorem. If $\sum_{j=1}^{n} \lambda_j^* > 1$ for all alternative solutions to (2), then it follows from part (a) of the theorem that $h_o^C < h_o^B$ and so $h_o^S < 1$. Let, if possible, $h_o^F < h_o^B$. Let $\lambda_j^{*(1)}$ be an optimal solution to (2) with $\alpha_1 = \sum_{j=1}^{n} \lambda_j^{*(1)} > 1$, and let $\lambda_j^{*(2)}$ be an optimal solution to (4) with $\alpha_2 = \sum_{j=1}^{n} \lambda_j^{*(2)} < 1$. Define

$$\lambda_j^{*(3)} = (1 - \alpha_3)\lambda_j^{*(1)} + \alpha_3 \lambda_j^{*(2)}, \quad j = 1, \dots, n,$$

where $0 < \alpha_3 = (\alpha_1 - 1)/(\alpha_1 - \alpha_2) < 1$. Then,

$$\sum_{j=1}^{n} \lambda_j^{*(3)} = (1 - \alpha_3)\alpha_1 + \alpha_3 \alpha_2 = 1,$$

$$\sum_{j=1}^{n} \lambda_j^{*(3)} Y_j = \sum_{j=1}^{n} \left[(1 - \alpha_3)\lambda_j^{*(1)} + \alpha_3 \lambda_j^{*(2)} \right] Y_j$$
$$\ge (1 - \alpha_3) Y_0 + \alpha_3 Y_0 = Y_0,$$

and

$$\sum_{j=1}^{n} \lambda_j^{*(3)} X_j \le (1-\alpha_3) h_o^C X_0 + \lambda_3 h_o^F X_0$$
$$= \left[(1-\alpha_3) h_o^C + \alpha_3 h_o^F\right] X_o.$$

Therefore,

$$h_o^B \le (1-\alpha_3) h_o^C + \alpha_3 h_o^F < (1-\alpha_3) h_o^B + \alpha_3 h_o^B = h_o^B,$$

which is a contradiction. Hence we conclude that it is not possible to have $h_o^F < h_o^B$ in this case. Thus, $h_o^F = h_o^B$.

The converse implication (the 'only if' part) for parts (b) and (c) follows immediately because the conditions specified in the theorem are mutually exclusive and exhaustive. □

3. Alternate optima and non-zero slack

Banker, Bardhan and Cooper (1996) show how to identify returns-to-scale possibilities from the optimal value for u_o in (3) without having to examine all alternate optima when the BCC model is used. We now similarly eliminate the need for examining all alternate optima when implementing the 'CCR Returns to Scale Theorem' of Banker and Thrall, as given in Section 2 above.

To start we suppose that an optimal solution with value h_o^C has been obtained for (2) with $\sum_{j=1}^{n} \lambda_j^* < 1$. To check on alternate optima possibilities we replace (2) with

$$\max \sum_{j=1}^{n} \hat{\lambda}_j + \varepsilon \left(\sum_{j=1}^{n} \hat{s}_i^- + \sum_{j=1}^{n} \hat{s}_r^+ \right) \tag{5}$$

subject to

$$\sum_{j=1}^{n} X_j \hat{\lambda}_j + \hat{s}^- = h_o^C X_o,$$
$$\sum_{j=1}^{n} Y_j \hat{\lambda}_j - \hat{s}^+ = Y_o,$$
$$\sum_{j=1}^{n} \hat{\lambda}_j \le 1,$$

where the components in the slack vectors $\hat{s}^-$ and $\hat{s}^+$, as well as the components of the vector $\hat{\lambda}$, are constrained to be non-negative. The symbol $\varepsilon > 0$ refers to a 'non-Archimedean' element which is smaller than *any* positive real number and, to avoid having to specify ε explicitly, DEA computer codes [6] generally utilize a two-stage process in which the sum of the slacks (as parenthesized in (5)) is maximized while fixing h_o^C at its optimal value. Now we note that the usual formulas for the 'CCR projections' give

$$\hat{x}_{io} = h_o^C x_{io} - \hat{s}_i^{-*} \le x_{io}, \quad i = 1,\ldots,m,$$
$$\hat{y}_{ro} = y_{ro} + s_r^{+*} \ge y_{ro}, \quad r = 1,\ldots,s, \tag{6}$$

where the $\hat{s}_i^{-*}$ and $\hat{s}_i^{+*}$ with non-negative values respectively represent optimal components of the slack vectors $\hat{s}_i^-$ and $\hat{s}_i^+$ in (5). As shown in Charnes, Cooper and Rhodes (1978), the $\hat{x}_{io}$ and $\hat{y}_{ro}$ in (6) are coordinates of points on the efficiency frontier. Because the values of $\sum_{j=1}^{n} \hat{\lambda}_j^*$ are maximal, the following theorem is immediate:

Theorem. *Given the existence of an optimal solution with* $\sum_{j=1}^{n} \lambda_j^* < 1$ *in model* (2), *returns to scale at* (X_o, Y_o) *are constant if and only if* $\sum_{j=1}^{n} \hat{\lambda}_j^* = 1$ *and returns to scale are increasing if and only if* $\sum_{j=1}^{n} \hat{\lambda}_j^* < 1$ *in* (5).

We are here restricting attention to solutions of (5) with $\sum_{j=1}^{n} \hat{\lambda}_j^* \le 1$. The example we provide below will show how to treat situations in which h_O^C is associated with solutions of (2) that have values $\sum_{j=1}^{n} \hat{\lambda}_j^* > 1$. Before proceeding to these examples, however, we introduce the following projection formulas from Banker and Morey (1986),

$$\hat{x}_{io}^* = \frac{h_o^C x_{io} - \hat{s}_i^{-*}}{\sum_{j=1}^{n} \hat{\lambda}_j^*}, \quad i = 1,\ldots,m,$$
$$\hat{y}_{ro}^* = \frac{y_{ro} + \hat{s}_r^{+*}}{\sum_{j=1}^{n} \hat{\lambda}_j^*}, \quad r = 1,\ldots,s, \tag{7}$$

[6] See Arnold et al. (1994) for general treatments and theorems underlying the use of such non-Archimedean concepts in DEA.

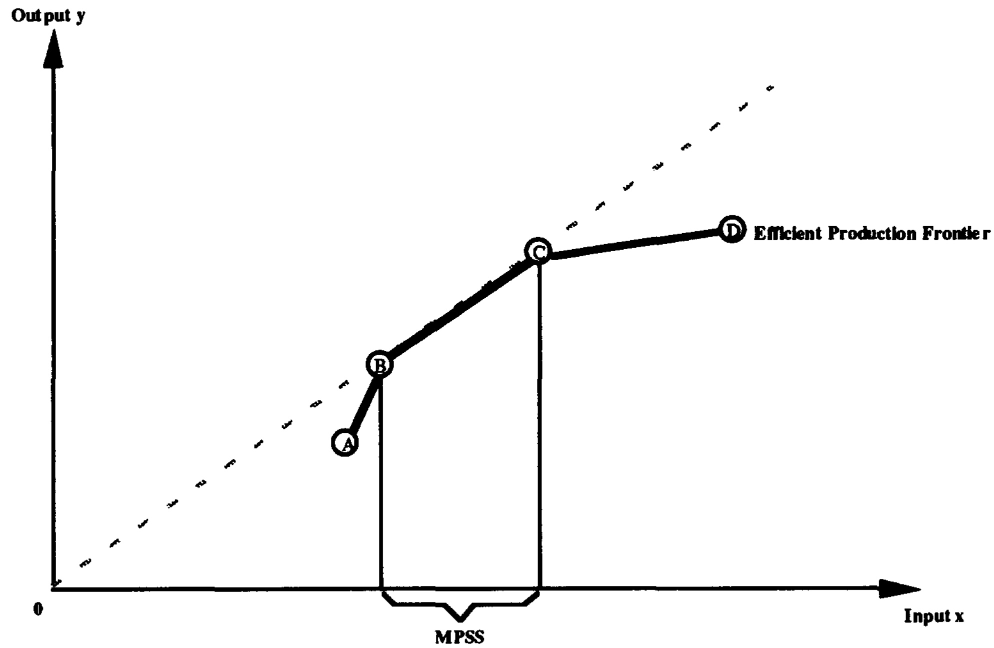

Fig. 1. Linear dependency and MPSS.

where the $\sum_{j=1}^{n}\hat{\lambda}_j^*$ and the $\hat{s}_i^{-*}$ and $\hat{s}_r^{+*}$ refer to optimal solutions for (5). The projection operators (6) and (7) used together will help to elucidate what is happening [7].

For illustration, we assign the following coordinate values to the points in Fig. 1:

$$\mathrm{A}=(1,1),\ \mathrm{B}=(\tfrac{3}{2},2),\ \mathrm{C}=(3,4),\ \mathrm{D}=(4,5),\ \mathrm{E}=(4,4\tfrac{1}{2}) \quad (8)$$

with the first component an input and the second an output. We have added $\mathrm{E}=(4,4\frac{1}{2})$ to the points in Fig. 1 to represent a DMU which, unlike the others, is not on either the CCR or the BCC efficiency frontier.

For A, which is BCC but not CCR efficient in Fig. 1, we substitute from (8) into (2) and write

$$h_o^C = \min h \quad (9)$$

subject to

$$h \geq \lambda_A + \tfrac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E,$$
$$1 \leq \lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \tfrac{9}{2}\lambda_E,$$
$$0 \leq \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E.$$

This problem has $h_o^C = \frac{3}{4}$ with alternate optima represented by $\lambda_B^* = \frac{1}{2}$ or $\lambda_C^* = \frac{1}{4}$ and all other $\lambda^* = 0$. In either case, we have $\sum_{j=1}^{n}\lambda_j^* < 1$ and so we utilize (5) to obtain

$$\max \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E + \varepsilon(\hat{s}^- + \hat{s}^+) \quad (10)$$

subject to

$$\tfrac{3}{4} \geq \hat{\lambda}_A + \tfrac{3}{2}\hat{\lambda}_B + 3\hat{\lambda}_C + 4\hat{\lambda}_D + 4\hat{\lambda}_E + \hat{s}^-,$$
$$1 \leq \hat{\lambda}_A + 2\hat{\lambda}_B + 4\hat{\lambda}_C + 5\hat{\lambda}_D + \tfrac{9}{2}\hat{\lambda}_E + \hat{s}^+,$$
$$1 \geq \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E,$$

so that, here, $\sum_{j=1}^{n}\hat{\lambda}_j \equiv \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E$ with all $\hat{\lambda}$ non-negative gives $\hat{\lambda}_B^* = \frac{1}{2}$ and all other variables zero – from which it follows that increasing returns to scale prevails for A. To see how this relates to the MPSS (Most Productive Scale Size) region of Fig. 1, we use (7) to obtain

$$\hat{x}_{io}^* = \frac{\frac{3}{4} - \hat{s}^{-*}}{\frac{1}{2}} = \frac{3}{2}, \qquad \hat{y}_{ro}^* = \frac{1 + \hat{s}^{+*}}{\frac{1}{2}} = 2 \quad (11)$$

and observe that this projects into B in Fig. 1 with coordinates given in (8) – which is MPSS in the sense of Banker (1984). Although it is not neces-

[7] See Banker et al. (1995b) for a detailed treatment of projection operators in DEA.

sary to do so, we can use the alternate optimum with $\lambda_C^* = \frac{1}{4}$ for $\Sigma_{j=1}^n \hat{\lambda}_j^*$ and obtain

$$\hat{x}_o^* = \tfrac{3}{4}/\tfrac{1}{4} = 3, \qquad \hat{y}_o^* = 1/\tfrac{1}{4} = 4, \tag{12}$$

which projects into C – also MPSS – and, of course, convex combinations of these alternate optima may be used to obtain other points between B and C which are *all* MPSS.

To evaluate E in (8) we alter the terms on the left in (9) to obtain

$$h_o^C = \min h \tag{13}$$

subject to

$$4h \geq \lambda_A + \tfrac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E,$$

$$\tfrac{9}{2} \leq \lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \tfrac{9}{2}\lambda_E,$$

$$0 \leq \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E.$$

We again have alternate optima with, now, $h_o^C = \frac{27}{32}$ for $\lambda_B^* = \frac{9}{4}$ or $\lambda_C^* = \frac{9}{8}$ and all other $\lambda^* = 0$. Here we have $\Sigma_{j=1}^n \lambda_j^* > 1$ in both cases. So, proceeding in an obvious modification of (5), we reorient the last constraint and the objective to obtain

$$\min\left(\hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E\right) - \varepsilon(\hat{s}^- + \hat{s}^+) \tag{14}$$

subject to

$$\tfrac{27}{8} = \hat{\lambda}_A + \tfrac{3}{2}\hat{\lambda}_B + 3\hat{\lambda}_C + 4\hat{\lambda}_D + 4\hat{\lambda}_E + \hat{s}^-,$$

$$\tfrac{9}{2} = \hat{\lambda}_A + 2\hat{\lambda}_B + 4\hat{\lambda}_C + 5\hat{\lambda}_D + \tfrac{9}{2}\hat{\lambda}_E - \hat{s}^+,$$

$$1 \leq \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E,$$

$$0 \leq \hat{\lambda}_A, \hat{\lambda}_B \hat{\lambda}_C, \hat{\lambda}_D, \hat{\lambda}_E.$$

This has its optimum at $\hat{\lambda}_C^* = \frac{9}{8}$ with all other variables equal to zero and so, in conformance with the Theorem for CCR models given in Section 2, we conclude that decreasing returns prevail at the point on the frontier to which this solution is projected by (6).

Again for further insight, we use (7) to obtain

$$\hat{x}_0^* = \frac{27}{8}/\frac{9}{8} = 3, \qquad \hat{y}_0^* = \frac{9}{2}/\frac{9}{8} = 4, \tag{15}$$

the coordinates for C in Fig. 1. Similarly an application to $\lambda_B^* = \frac{9}{4}$, the alternate optimum to (13), gives

$$\hat{x}_0^* = \tfrac{27}{8}/\tfrac{9}{4} = \tfrac{3}{2}, \qquad \hat{y}_0^* = \tfrac{9}{2}/\tfrac{9}{4} = 2, \tag{16}$$

which are the coordinates for B.

We now note that C and B serve as the basis members with positive coefficients in all of our optima and that these are points in the interval that is common to both the BCC and CCR efficiency frontiers. All points in that interval formed by convex combinations of B and C are MPSS and these are the only points that are both CCR and BCC efficient.

Using the above examples for guidance we can see that the situation is general from the following considerations. First, as shown in Ahn, Charnes and Cooper (1989) a point which is found to be efficient for the CCR model will also be efficient for the BCC model (but the converse is not necessarily true) so all such CCR efficient points lie in the intersection of the CCR and BCC efficiency frontiers. Second, as shown in Charnes et al. (1978) a DMU can be part of an optimal basis with a positive coefficient only if it is efficient. Hence all members of an optimal basis with positive coefficients lie in the intersection of the CCR and BCC efficiency frontiers. The rest of what is needed follows from the other properties of the efficiency frontiers of both the BCC and CCR models – viz., such frontiers are concave, monotonically increasing and continuous. In short the associated frontier function is 'isotonic' in the manner described in Charnes et al. (1985).

The following mathematical formulation shows what is happening analytically. When (2) and (5) are used to evaluate DMU_o with input vector X_o and output vector Y_o we can write the optimal solution as

$$P_E = \begin{pmatrix} h_o^C X_o - \hat{s}^{+*} \\ Y_o + \hat{s}^{-*} \end{pmatrix} = \sum_{j=1}^n P_j \hat{\lambda}_j^*. \tag{17}$$

Here P_E is a point on the efficient production frontier. It is the point used to evaluate DMU_o with coordinate values the same as those obtained by using the CCR projection operators defined in (6). In short, the $\hat{\lambda}_j^*$ in (17) effect this same projection and so our returns-to-scale characterizations are also obtained from these $\hat{\lambda}_j^*$ values for points on the efficient production frontier. Because P_E is on the CCR efficiency frontier, it has constant returns-to-scale. Hence the Banker–Morey projection in (7) may be used to bring P_E into MPSS, when desired, with no need to adjust the observed mix to the efficiency proportions required at MPSS because this has already been accomplished by (17). In any case, the value of $\sum_{j=1}^n \hat{\lambda}_j^*$ serves to relate the scale at which P_E operates to what is required for MPSS with $\sum_{j=1}^n \hat{\lambda}_j^* < 1$ when expansion is indicated and $\sum_{j=1}^n \hat{\lambda}_j^* > 1$ when contraction is needed to achieve MPSS.

We should perhaps note that there may be more than one P_E if alternate optima are present in (5). This presents no problem because the same optimal value for the sum of the $\hat{\lambda}$s will be applicable for all such alternate optima. However, as noted in Banker and Thrall (1992), the situation for returns to scale may change if an 'output-oriented' version of the CCR model is used. In important applications, it may therefore be a good idea to use both the 'input' and the 'output oriented' versions of these models and choose between them according to whether output augmentation or input conservation is to be emphasized.

4. Conclusion

A variety of strategies are evidently opened by the preceding developments which can be used in addressing returns-to-scale considerations. One can, of course, use the two-stage approach described in Färe, Grosskopf and Lovell (1985) with its accompanying 'strong' or 'weak disposal' assumptions. However, proceeding along the lines described in the preceding section one can dispense with these assumptions and one need not assume that efficiency has first been achieved to apply the CCR Returns to Scale Theorem given in Section 2. There is yet another (third) approach via the BCC model described in Banker and Thrall (1992) and modified in Banker, Bardhan and Cooper (1996) to avoid the need for (a) examining all alternate optima and (b) assuming that efficiency has first been achieved in order to use the BCC Returns to Scale Theorem (also given in Section 2, immediately following (3.0), above).

Examples like the following suggest some of the possibilities.

Example 1. Arnold, Bardhan and Cooper (1995) elected the CCR approach in a study of Texas schools because not all members of the study consortium were familiar with DEA. Hence it seemed best to avoid issues involved in distinguishing between technical and scale efficiencies when putting the study under way. A study of returns-to-scale efficiencies could then be brought forth, if desired, only at a later stage of this study when others would be more familiar with DEA and its uses [8].

Example 2. Banker, Conrad and Strauss (1986) redid a study of North Carolina hospitals in order to reexamine results from the earlier study's use of statistical regressions which found that constant returns to scale prevailed. A use of the BCC model for this purpose by Banker et al. made it possible to take advantage of the distinctions between technical efficiency, as treated in the primal (1), and returns-to-scale inefficiency, as treated in the dual (3), because this separation could help to locate sources of differences in the results obtained from these DEA and statistical regression approaches.

These are only examples of the possibilities. As should now be clear, *either* the CCR or BCC model may be used when interest is only on as whether returns to scale are increasing, decreasing or constant in one or more DMUs. Numerical values, such as rates of return or scale elasticities, may also be obtained as described in Banker and Thrall (1992) and as interpreted and elaborated in Cooper, Thompson and Thrall (1996).

[8] We might also note that an examination conducted at a later stage of this study found that returns-to-scale inefficiencies were small (as was conjectured) relative to much greater amounts of technical inefficiency identified in the first study stage for these schools (Bardhan, 1995).

Acknowledgements

Acknowledgment is due to R.M. Thrall and an anonymous referee for their constructive comments and suggestions. Support from the IC^2 Institute of the University of Texas at Austin is gratefully acknowledged.

References

Ahn, T., Charnes, A., and Cooper, W.W. (1989), "A note on the efficiency characterizations obtained in different DEA models", *Socio-Economic Planning Sciences.*

Arnold, V., Bardhan, I., and Cooper, W.W. (1996), "A two-stage DEA approach for identifying and rewarding efficiency in Texas public schools", *Impact: Essays in Honor of G. Kozmetsky*, Kluwer, Norwell, MA (forthcoming)

Arnold, V., Bardhan, I., Cooper, W.W., and Gallegos, A. (1995), "Primal and dual optimality in IDEAS (Integrated Data Envelopment Analysis Systems) and related computer codes", in: J. Aronson and S. Zionts (eds.), *Operations Research: Models, Methods and Applications. A Volume in Honor of G.L. Thompson*, Quorum Books, New York.

Banker, R.D. (1984), "Estimating most productive scale size using Data Envelopment Analysis", *European Journal of Operational Research* 17, 35–44.

Banker, R.D., and Morey, R. (1986), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research* 34, 513–521.

Banker, R.D., and Thrall, R.M. (1992), "Estimation of Returns to Scale using Data Envelopment Analysis", *European Journal of Operational Research* 62, 74–84.

Banker, R.D., Charnes, A., and Cooper, W.W. (1984), "Models for the estimation of technical and scale inefficiencies in Data Envelopment Analysis", *Management Science* 30, 1078–1092.

Banker, R.D., Conrad, R.F., and Strauss, R.P., (1986), "A comparative application of DEA and translog methods: An illustrative study of hospital production", *Management Science* 36, 30–34.

Banker, R.D., Charnes, A., Cooper, W.W., Swarts, J., and Thomas, D. (1989), "An introduction to Data Envelopment Analysis with some of their models and their uses", *Research in Governmental and Non Profit Accounting* 5, 125–163.

Banker, R.D., Chang, H., and Cooper, W.W. (1995), "Simulation studies of efficiency and returns to scale and misspecification with nonlinear functions in DEA", *The Annals of Operations Research* (W. Cooper, R. Thompson, R. Thrall, eds.).

Banker, R.D., Bardhan, I., and Cooper, W.W., (1996), "A note on returns of scale in DEA", *European Journal of Operational Research* 88, 583–585.

Bardhan, I. (1995), "DEA and stochastic frontier regression studies of public school performances in Texas", Ph.D. Thesis, University of Texas, Graduate School of Business, Austin, TX.

Byrnes, P., Färe, R., and Grosskopf, S. (1984), "Measuring productive efficiency: An application to Illinois strip mines", *Management Science* 30, 671–680.

Chang, K., and Guh, Y. (1991), "Linear production functions and the Data Envelopment Analysis", *European Journal of Operational Research* 52, 215–223.

Charnes, A., Cooper, W.W., and Rhodes, E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research* 2, 429–441.

Charnes, A., Cooper, W.W., and Rhodes, E. (1981), "Data Envelopment Analysis as an approach for evaluating program and managerial efficiency —- with an illustrative application to the Program Follow Through Experiment in U.S. public school education", *Management Science* 27, 668–697.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L., and Stutz, J. (1985), "Foundations of Data Envelopment Analysis and Pareto-Koopmans empirical production functions", *Journal of Econometrics* 30, 91–107.

Cooper, W.W., Thompson, R.G., and Thrall, R.M. (1996), "Introduction: Extensions and new developments in DEA", *Annals of Operations Research* (to appear).

Färe, R., and Grosskopf, S. (1994), "Estimation of returns to scale using Data Envelopment Analysis", *European Journal of Operational Research* 79, 379–382.

Färe, R., Grosskopf, S., and Lovell, C.A.K. (1985), *The Measurement of Efficiency of Production*, Kluwer Nijhoff, Boston, MA..

Ganley, J.A., and Cubbin, J.S. (1992), *Public Sector Efficiency Measurement: Applications of Data Envelopment Analysis*, Elsevier, Amsterdam.

Koopmans, T.C. (1951), "Analysis of production as an efficient combination of activities", in: T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Wiley, New York.

Thrall, R.M. (1996), "Duality, classification and slacks in DEA", *Annals of Operations Research* (W. Cooper, R. Thompson and R. Thrall, eds.) (to appear).

Koopmans, T.C. (1957), *Three Essays on the State of Economic Science*, McGraw-Hill, New York.

Zhu, J., and Shen, Z. (1992), "Most productive scale size and returns to Scale", Working Paper, Economics and Management School, Southeast University, Nanjing, People's Republic of China.

Zhu, J., and Shen, Z. (1993), "A discussion of testing DMUs' returns to scale", Working Paper, Economics and Management School, Southeast University, Nanjing, People's Republic of China.