ELSEVIER

European Journal of Operational Research 103 (1997) 28–37

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Estimating returns to scale in DEA

Boaz Golany [a,b,*], Gang Yu [c,d]

[a] *IC² Institute, University of Texas at Austin, Austin, TX 78705, USA*

[b] *Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 32 000, Israel*

[c] *Department of Management Science and Information Systems, College of Business Administration, University of Texas at Austin, Austin, TX 78712, USA*

[d] *Center for Cybernetic Studies, University of Texas at Austin, Austin, TX 78712, USA*

Received 24 May 1994; accepted 19 July 1996

Abstract

This paper addresses issues of returns to scale in Data Envelopment Analysis. Starting with the model developed by Banker, but avoiding Banker's conclusions on returns-to-scale, the paper shows how two close variants (inputs and outputs oriented) of the Banker–Charnes–Cooper model can be used to provide precise estimates of returns to scale. The estimation of returns to scale for each unit is done by testing the existence of solutions in four regions defined in the neighborhood of the analyzed unit. Numerical examples and graphs are used to illustrate the proposed procedures. © 1997 Elsevier Science B.V.

Keywords: Data Envelopment Analysis (DEA); Returns to scale (RTS)

1. Introduction

Data Envelopment Analysis (DEA) is a methodology created to evaluate the relative efficiency of Decision Making Units (DMUs) which use similar types of (multiple) resources to produce similar kinds of (multiple) outputs. The methodology was introduced by Charnes, Cooper and Rhodes (CCR) (1978) and has been widely applied since then in different production situations in the public as well as in the private sectors (see Seiford, 1991). The topic of relative efficiency is closely related to classical notions of productivity (see, e.g., Brinkerhoff and Dressler, 1990).

The productivity literature discusses productivity improvements in five basic ways:

- producing the same output(s) while consuming less resources (inputs);
- producing more output(s) without changing the level of resources used;
- producing more outputs with fewer inputs;
- a larger increase in the outputs for an increase in the inputs; and
- a smaller reduction in the outputs for a decrease in inputs consumption.

In the DEA terminology, the first three of the productivity improvement directions listed above are associated with 'technical efficiency' while the latter two fall into the category of 'scale efficiency'. The need to untangle these two sources of potential inefficiencies was recognized by economists working in

* Corresponding author.

PII S0377-2217(96)00259-7

the area of production efficiency (see, e.g., Färe, Grosskopf and Lovell, 1985, Section 8). The estimation of returns to scale (RTS) in DEA was first investigated in Banker (1984) and Banker, Charnes and Cooper (BCC) (1984). Both of these studies presented a modification to the CCR model (by adding a convexity constraint) and proposed a technique for estimating RTS. However, this technique was based on the assumption of unique optimal solution to the linear programming formulation of the BCC model. Unfortunately, in almost all of DEA applications one encounters alternate optima for some DMUs. Chang and Guh (1991) highlighted the problem using counter examples to demonstrate it. Banker and Thrall (BT) (1992) acknowledged the problem with the earlier technique of measuring RTS and suggested ways to handle the general case (i.e., non-unique optimal solutions). The modified technique of BT requires knowledge about all optimal solutions – a task which is handled by solving two auxiliary linear programming models for each DMU being evaluated. However, even after solving these auxiliary programs the modified technique does not provide unambiguous detection of RTS in situations where u_0, Banker's original RTS indicator, has a range of both negative and positive optimal values. [1] In these cases, BT define the RTS as constant. But, as shown here, further investigation of these situations may reveal a more precise information on different RTS in different directions around the DMU being evaluated.

Also, the Banker–Thrall technique is applicable only for technically efficient DMUs. While it is true that technically inefficient DMUs must first eliminate the waste in their operations (redundant levels of resources and/or insufficient levels of outputs), it is still interesting to find out, as suggested in the present paper, the directions of RTS that might be associated with the region in which these units operate.

Another point which was left unnoticed in most of the previous RTS work is the fact that the RTS is a local phenomenon. This point is discussed at some length in Banker et al. (1989, p. 145) where the authors argue that the RTS found through the BCC model (putting aside the deficiency related to the non-uniqueness assumption) hold only in DMU_0's current position. The implication of this argument, which was never explored up till now, is that the identification of RTS for a particular DMU without executing an additional sensitivity analysis in its immediate neighborhood is meaningless.

This paper develops a simple procedure, based on two linear programming variants of the BCC model (input and output oriented formulations), to estimate the RTS. Unlike previous techniques, the estimation procedure proposed here does not seek exact RTS magnitudes. Rather, its objective is to provide an accurate classification of RTS into constant, increasing or decreasing rates. Further, the proposed technique offers at least a partial remedy to the identification problems associated with the local nature of the RTS property by performing a sensitivity analysis based on observing the RTS behavior in two specified directions around the DMU being analyzed.

2. Model development

The data assumed for the model consist of input–output observations describing the productive performance of n DMUs ($j = 1, \ldots, n$). For each DMU, a vector X_j of m inputs $(X_{1j}, \ldots, X_{mj})$ and a vector Y_j of s outputs $(Y_{1j}, \ldots, Y_{sj})$ are observed. To define the various terms used in the sequel we follow Charnes et al. (1985) in characterizing the production possibility set T as

$$T = \left\{ (X, Y) \mid X \geqslant \sum_{j=1}^{n} X_j \cdot \mu_j, \quad Y \leqslant \sum_{j=1}^{n} Y_j \cdot \mu_j, \right.$$
$$\left. \sum_{j=1}^{n} \mu_j = 1, \ \mu_j \geqslant 0 \right\}. \tag{1}$$

Specifically, T is the convex hull of the observed points (X_j, Y_j), $j = 1, \ldots, n$. Consequently (Charnes et al., Theorem 2, p.95), the empirical frontier function that lies above T is a concave, piece-wise linear function. A particular DMU_0 is said to be technically efficient if no convex combination of other DMUs can be found with all outputs greater than or equal to

[1] See the Appendix for a simple example that demonstrates the ambiguity of the results given by this modified technique.

the outputs of DMU_0 and all inputs smaller than or equal to the inputs of DMU_0 with at least one strict inequality. In other words, DMU_0 is technically efficient if $\{\mu_0 = 1, S_r^+ = S_i^- = 0, \forall r, i\}$ is an optimal solution to the following program (known as the Additive DEA model).

$$\text{Max} \quad \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \tag{2}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \mu_j - s_r^+ = Y_{r0}, \quad r = 1, \ldots, s,$$

$$\sum_{j=1}^{n} X_{ij} \cdot \mu_j + s_i^- = X_{i0}, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_j = 1,$$

$$\mu_j,\ s_i^-,\ s_r^+ \geqslant 0,$$

where S_r^+, S_i^- represent output and input slacks, respectively.

If a DMU_0 is technically inefficient, it can be projected to an efficient position (X_p, Y_p) by

$$X_{ip} = X_{i0} - S_i^{-*}, \quad i = 1, \ldots, m,$$

$$Y_{rp} = Y_{r0} + S_r^{+*}, \quad r = 1, \ldots, s,$$

where the asterisks represent (throughout the paper) optimal solution values.

Now, let α_0 be a proportional change (increase or decrease) in all the outputs of DMU_0 and β_0 be a proportional change in all its inputs. The set of all feasible proportional (radial) changes associated with DMU_0, $F(X_0, Y_0)$, can then be defined by

$$F(X_0, Y_0) = \left\{ (\alpha_0, \beta_0) \mid \beta_0 \cdot X_0 \geqslant \sum_{j=1}^{n} X_j \cdot \mu_j, \right.$$

$$\alpha_0 \cdot Y_0 \leqslant \sum_{j=1}^{n} Y_j \cdot \mu_j,$$

$$\left. \sum_{j=1}^{n} \mu_j = 1,\ \mu_j \geqslant 0 \right\}. \tag{3}$$

To find the best possible improvement in the productivity (i.e., technical and scale efficiency) of DMU_0

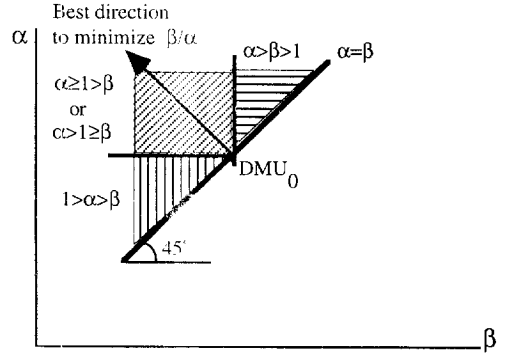

Fig. 1. Projection regions for DMU_0.

one would like to minimize β_0 while maximizing α_0. This can be achieved by either minimizing the ratio β_0/α_0, or maximizing its reciprocal α_0/β_0, subject to constraints that define the production possibilities set. The optimal values of α_0 and β_0 indicate the RTS associated with DMU_0. Graphically, Fig. 1 defines for DMU_0 four regions in the two dimensional space (α, β) where possible projection to another production possibility may be sought.

The projection regions are:

(i) $1 > \alpha_0 > \beta_0$ indicates decreasing RTS,

(ii) $\alpha_0 > \beta_0 > 1$ indicates increasing RTS,

(iii) $\alpha_0 = \beta_0$ indicates constant RTS.

(iv) $\alpha_0 > 1 \geqslant \beta_0$ or $\alpha_0 \geqslant 1 > \beta_0$ indicates technical inefficiency associated with DMU_0.

In region (i), both values are smaller than one and β_0^* is strictly smaller than α_0^*. This means that a feasible production possibility can be found that, compared with DMU_0, uses fewer inputs to produce less outputs but in doing so it contracts the inputs by a larger factor than the reduction in outputs, i.e., a decreasing returns to scale situation. Similar arguments explain regions (ii) and (iii). If projection is possible into region (iv), it means that DMU_0 has failed the test in (2) and is therefore technically inefficient.

When minimizing β_0/α_0 subject to the constraints defined by (3), $\alpha_0 = \beta_0 = \mu_0 = 1$ is always a feasible solution for any DMU_0. Hence, the optimal solution must always satisfy $\alpha_0^* \geqslant \beta_0^*$, resulting in all the projection regions in Fig. 1 being on or above the 45° line of $\alpha = \beta$. The theorem below identifies

the best direction to take (at $\alpha = \beta = 1$) in minimizing β_0/α_0.

Theorem 1. *The optimal direction minimizing β_0/α_0 is the normal to the line $\beta = \alpha$.*

Proof. Since RTS are defined as a local property, we only need to examine them in an ε-neighborhood of the point $(\alpha_0, \beta_0) = (1, 1)$. Assume that a new point (α', β') is obtained by an infinitesimal displacement ε from (α_0, β_0), i.e.,

$$(\alpha', \beta') = (1 + \varepsilon \cdot \sin\theta, 1 + \varepsilon \cdot \cos\theta),$$

where θ is the polar angle. The ratio we seek to minimize can now be written as

$$\frac{\beta'}{\alpha'} = \frac{1 + \varepsilon \cdot \cos\theta}{1 + \varepsilon \cdot \sin\theta} = 1 + \varepsilon \cdot [\cos\theta - \sin\theta] + O(\varepsilon^2) = f(\theta).$$

Ignoring the higher order term, $O(\varepsilon^2)$, $f(\theta)$ is minimized at $\theta^* = \frac{3}{4}\pi$. This can be checked by noting that

$$f'(\theta^*) = 0,$$

$$f''(\theta^*) = \sqrt{2} \cdot \varepsilon > 0. \qquad \square$$

The model development underlying definition (3) can be traced to Banker (1984) who established the link between it and the CCR model. Starting with the CCR model,

$$\text{Min} \quad h_0 - \varepsilon \cdot \left[\sum_{i=1}^{m} s_{i0}^- + \sum_{r=1}^{s} s_{r0}^+ \right] \tag{4}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \lambda_j - s_{r0}^+ = Y_{r0}, \quad r = 1, \ldots, s,$$

$$h_0 \cdot X_{i0} - \sum_{j=1}^{n} X_{ij} \cdot \lambda_j - s_{i0}^- = 0, \quad i = 1, \ldots, m,$$

$$\lambda_j,\ s_{i0}^-,\ s_{r0}^+ \geqslant 0.$$

Banker defined a new variable k_0 as

$$k_0 = \sum_{j=1}^{n} \lambda_j.$$

Since k_0 is always positive it can be used to divide through the two constraint groups of (4):

$$\text{Min} \quad h_0 - \varepsilon \cdot \left[\sum_{i=1}^{m} s_{i0}^- + \sum_{r=1}^{s} s_{r0}^+ \right] \tag{5}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \frac{\lambda_j}{k_0} - \frac{s_{r0}^+}{k_0} = \frac{1}{k_0} \cdot Y_{r0}, \quad r = 1, \ldots, s,$$

$$\frac{h_0}{k_0} \cdot X_{i0} - \sum_{j=1}^{n} X_{ij} \cdot \frac{\lambda_j}{k_0} - \frac{s_{i0}^-}{k_0} = 0, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \lambda_j / k_0 = 1,$$

$$\lambda_j,\ s_{i0}^-,\ s_{r0}^+ \geqslant 0.$$

Now, defining new variables α_0, β_0 and μ_j as

$$\mu_j = \frac{\lambda_j}{k_0}, \qquad \alpha_0 = \frac{1}{k_0}, \qquad \beta_0 = \frac{h_0}{k_0}$$

leads to the following mathematical program: [2]

$$\text{Min} \quad \frac{\beta_0}{\alpha_0} - \varepsilon \cdot \left[\sum_{i=1}^{m} s_{i0}^- + \sum_{r=1}^{s} s_{r0}^+ \right] \tag{6}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \mu_j - s_{r0}^+ = \alpha_0 \cdot Y_{r0}, \quad r = 1, \ldots, s,$$

$$\sum_{j=1}^{n} X_{ij} \cdot \mu_j + s_{i0}^- = \beta_0 \cdot X_{i0}, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_j = 1,$$

$$\mu_j,\ s_{i0}^-,\ s_{r0}^+ \geqslant 0.$$

Notice that by abuse of notation ε/k_0 is still denoted as ε because of the non-Archimedean character of this parameter.

Banker proceeded to interpret the value of k_0 as an indicator of RTS (Banker, 1984, p. 40, Corollary 1). This interpretation was based on the assumption

[2] Banker (1984, p. 40) wrote only the constraints of this program and omitted the objective function.

that k_0 has a unique optimal value. However, as Chang and Guh (1991) show, there can be alternate optima of k_0 with values above and below one for the same program. Thus, observing an optimal k_0 value in the CCR formulation does not tell us anything about RTS. Similarly, observing the optimal value of h_0 in the CCR model is not enough to learn the RTS direction. The same optimal value, say $h_0^* = 0.8$, can be obtained from $\beta_0^* = 1.2$ and $\alpha_0^* = 1.5$ (indicating increasing RTS) or from $\beta_0^* = 0.64$ and $\alpha_0^* = 0.8$ (indicating decreasing RTS).

Thus, one can not rely on the CCR model (4) to obtain correct estimates of RTS. Instead, we focus our attention again on (6). Rather than solving this non-linear model we propose to solve two linear variants of it. The purpose of these programs is to test in which of the four regions defined above do feasible solutions exist within a small distance from DMU_0. As proven in Theorem 2, the existence of a solution in a particular region provides sufficient information to classify the RTS around DMU_0.

First, we set α_0 equal to one plus a small arbitrary number (i.e., $\alpha_0 = 1 + \delta$, $\delta > 0$). This transforms the model to

$$\text{Min} \quad \beta_0 - \varepsilon \cdot \left[\sum_{i=1}^{m} s_{i0}^- + \sum_{r=1}^{s} s_{r0}^+ \right] \tag{7}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \mu_j - s_{r0}^+ = (1 + \delta) \cdot Y_{r0}, \quad r = 1, \ldots, s,$$

$$\sum_{j=1}^{n} X_{ij} \cdot \mu_j + s_{i0}^- = \beta_0 \cdot X_{i0}, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_j = 1,$$

$$\mu_j,\ s_{i0}^-,\ s_{r0}^+ \geq 0.$$

Second, we set β_0 to be equal to one minus a small number (i.e., $\beta_0 = 1 - \delta$, $\delta > 0$). This yields the following model:

$$\text{Max} \quad \alpha_0 + \varepsilon \cdot \left[\sum_{i=1}^{m} s_{i0}^- + \sum_{r=1}^{s} s_{r0}^+ \right] \tag{8}$$

s.t.

$$\sum_{j=1}^{n} Y_{rj} \cdot \mu_j - s_{r0}^+ = \alpha_0 \cdot Y_{r0}, \quad r = 1, \ldots, s,$$

$$\sum_{j=1}^{n} X_{ij} \cdot \mu_j + s_{i0}^- = (1 - \delta) \cdot X_{i0}, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_j = 1,$$

$$\mu_j,\ s_{i0}^-,\ s_{r0}^+ \geq 0.$$

We now propose a procedure for estimating RTS on the basis of programs (7)–(8):

A procedure for estimating RTS

Step 1. Solve (7) to determine the RTS to the 'right' of DMU_0:

Step 1(i). $1 + \delta > \beta_0^* > 1 \Rightarrow$ increasing RTS.

Step 1(ii). $1 \geq \beta_0^* \Rightarrow$ DMU_0 is inefficient.

Step 1(iii). $1 + \delta = \beta_0^* \Rightarrow$ constant RTS.

Step 1(iv). $1 + \delta < \beta_0^* \Rightarrow$ decreasing RTS.

Step 1(v). No feasible solution $\Rightarrow$ there is no data to determine the RTS to the 'right' of DMU_0.

Step 2. Solve (8) to determine the RTS to the 'left' of DMU_0:

Step 2(i). $1 > \alpha_0^* > 1 - \delta \Rightarrow$ decreasing RTS.

Step 2(ii). $\alpha_0^* \geq 1 \Rightarrow$ DMU_0 is inefficient.

Step 2(iii). $1 - \delta = \alpha_0^* \Rightarrow$ constant RTS.

Step 2(iv). $\alpha_0^* < 1 - \delta \Rightarrow$ increasing RTS.

Step 2(v). No feasible solution $\Rightarrow$ there is no data to determine the RTS to the 'left' of DMU_0.

Theorem 2. *Assuming that the empirical frontier is a concave, piece-wise linear function the procedure outlined above yields correct estimates of* RTS.

Proof. Fig. 2 will facilitate the proof by presenting the location of DMU_0 in the two-dimensional space (α_0, β_0).

Fig. 2 depicts a situation in which DMU_0 lies on a decreasing RTS piece of the frontier. Consequently, when program (8) is solved and an optimal solution is found with $1 - \delta < \alpha_0^* < 1$ (both inequalities strictly held), it means that:

- no feasible solution exists in region (iv) since otherwise it would have been better than the solution found now;

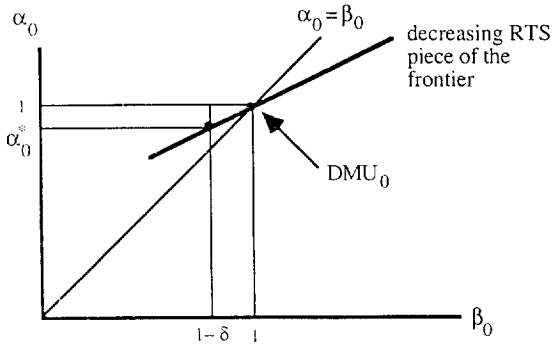

Fig. 2. Position of DMU_0 in (α_0, β_0) space.

● no feasible solution exists in region (i) since the existence of a convex combination of other DMUs in region (ii), established by the current solution, indicates that DMU_0 lies on a decreasing RTS piece of the frontier and the piece-wise concavity assumption precludes the possibility of an increasing RTS piece to the 'right' of a decreasing RTS piece.

When $1-\delta > \alpha_0^*$ we know that there exists a convex combination of other DMUs under the 45° line $\alpha = \beta$, i.e., there exists an increasing RTS piece of the frontier to the left of DMU_0. Similar arguments may be used to prove the other cases given in the theorem. □

The procedure need not always be solved in its entirety and the order of the two steps is insignificant. When an inefficient unit is detected through (7) there is no need to confirm this finding through (8). For efficient DMUs, however, their location on edges of the frontier make them likely to be associated with more than one type of RTS. Hence, it is recommended to solve both (7) and (8) for these units. The next two sub-sections focus on specific analysis that can be affected within the context of the proposed RTS procedure to efficient and inefficient DMUs.

2.1. Treating inefficient DMUs

When analyzing an inefficient DMU_0 one may find that the unit can be associated with more than one facet of the empirical frontier function, depending on whether output or input improvements are deemed to be more important. Hence, such unit can be associated with more than one estimate of RTS. The latter case is illustrated in Fig. 3.

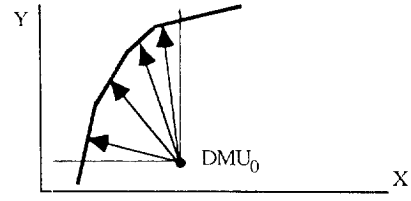

Fig. 3. Projections of an inefficient DMU_0 to segments of the frontier with different RTS.

To force a correspondence to a single piece of the frontier we construct two projected DMUs. The first, DMU_{0L}, is generated by running the BCC input oriented model, the second, DMU_{0R}, through the BCC output oriented model. This procedure is illustrated in Fig. 4.

If at the end of the procedure suggested above both projected DMUs are related to the same RTS then, in spite of DMU_0's technical inefficiency, we can claim that DMU_0 is associated with this RTS. However, if the situation resembles what is seen in Fig. 3, the two projected DMUs will be associated with different RTS and we will be unable to determine exact RTS for DMU_0.

2.2. Treating efficient DMUs

The task of estimating the RTS for technically efficient DMUs is made more complicated in situations where such DMUs are positioned on an edge of the frontier separating pieces with different RTS (e.g., as can be seen with the unnamed DMU on the frontier in Fig. 4). In fact, due to the complex geometry which characterize the multi-dimensional frontier, an efficient DMU may be located at the edge of many pieces with different RTS. As the task of testing in all possible directions is exceedingly

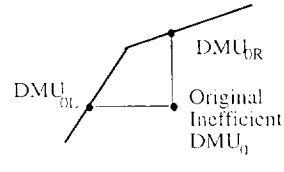

Fig. 4. Analysis of inefficient DMUs.

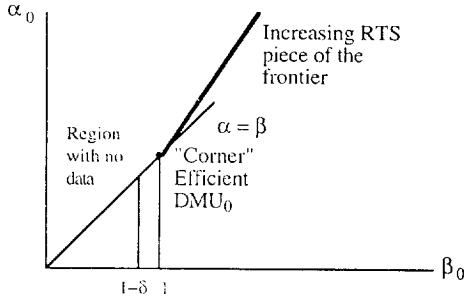

Fig. 5. No feasible solution to program (7).

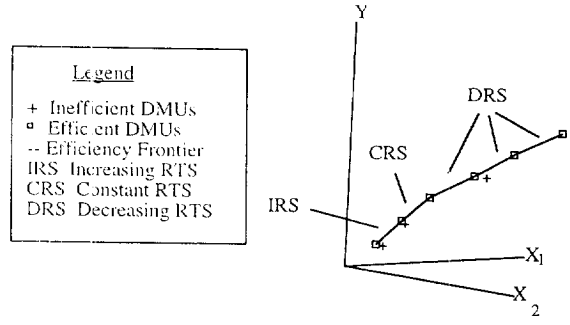

Fig. 6. A 3-D illustration of the numerical example.

difficult, we limit ourselves to testing in two directions.

The first test, 'on the right', is performed by raising all outputs by a small factor and the second test, 'on the left', is done by reducing all inputs by a small factor. If the DMU happens to be at a 'corner' of the frontier and we impose through (7) or (8) a shift outside the area which is under the frontier, there is no feasible solution to the relevant program. These situations are handled in Steps 1(v) and 2(v) of the procedure. Fig. 5 illustrates a situation with no feasible solution to program (7).

3. Numerical example

We consider a group of 9 DMUs using two inputs and producing a single output. The DMUs are depicted in Fig. 6 and their (X, Y) data along with the solutions to programs (7) and (8) are given in Table 1.

The solution of (7) for DMU_1 is

$\beta^* = 1.00008 < 1 + \delta$.

Hence, according to Step 1(i) in the RTS procedure, increasing RTS prevail to the right of DMU_1. Attempting to solve (8) for the same DMU results in no feasible solution (illustrating the situation depicted in Fig. 5). DMUs 2, 4 and 9 are found to be inefficient by both programs (7) and (8). As explained earlier, it was not necessary to solve both programs to obtain this result and it was done here simply to demonstrate the consistency between the two formulations. Solving (8) for DMU_3 reveals increasing RTS on its left – corresponding to the RTS found through (7) to the right of DMU_1 (refer to Figure 6 where this increasing RTS piece of the frontier is indicated).

Now, notice the changes in the direction of RTS that happen when we go across the efficient DMUs 3 and 5. Starting with an increasing RTS left of DMU_3, we move to a constant RTS to the right of it and to the left of DMU_5 and end with decreasing RTS on the right of DMU_5 and for all other DMUs whose inputs are larger than those of DMU_5.

Table 1
Solutions of (7) and (8) for all DMUs ($\delta = 0.0001$)

DMU	X_1	X_2	Y	β_0^* (7)	RTS on the right	α_0^* (8)	RTS on the left
1	1.1	1.1	1	1.00008	IRS	N/A	N/A
2	1.4	1.3	0.95	0.84615	inefficient	1.2864	inefficient
3	2	2	2	1.0001	CRS	0.99988	IRS
4	2.1	2.2	1.9	.90960	inefficient	1.1051	inefficient
5	3	3	3	1.000175	DRS	0.9999	CRS
6	4.2	5	4	1.000123	DRS	0.9999125	DRS
7	5.5	6.5	5	1.000153	DRS	0.999915	DRS
8	7	8.5	6	N/A	N/A	0.999922	DRS
9	5	5	3.9	.915136	inefficient	1.0622	inefficient

To provide further analysis of the RTS that can be associated with the inefficient DMUs (2, 4 and 9) we performed the treatment suggested in Section 2.1. First, each of the inefficient DMUs was projected twice to the efficient frontier – once using the input-oriented BCC model and the second time using the output oriented BCC model. The (X, Y) values of the projected units along with the solution of (7) for these units are given in Table 2.

As expected, in all three cases the two projected points for each DMU are located at distinctly different positions on the frontier. Repeating the analysis for all six projected points showed that:

- the projections for DMU_2 were both associated with IRS;
- the projections for DMU_9 were both associated with DRS;
- the projections for DMU_4 were associated with CRS and DRS.

Hence, although these are inefficient DMUs, we can now associate two of them (2 and 9) with a single RTS condition. The third, DMU_4, is in a similar situation to the efficient DMU_3 which can be associated with both CRS and DRS depending on the direction which is investigated.

4. Summary and concluding remarks

This paper proposes a simple technique, based on solving two variants of the BCC model to estimate the RTS of the units evaluated through DEA. The computational requirements for this estimation are minimal since the two formulations differ only slightly from existing models – thus making it easy to modify existing DEA codes to solve them.

An important advantage of this technique over previous ones is the fact that it provides sharper results for efficient DMUs and (at least partial) results for inefficient DMUs that were not addressed at all by previous works. However, one should bear in mind that the RTS estimates which the proposed technique (and, for that matter, any other technique) is offering are true only locally and only in the directions that were specified (here, to the 'right' and 'left' of the analyzed DMU). Experience with DEA shows that in many cases the facets that are found along the empirical frontier do not enjoy full dimensionality, i.e., they are determined by linear equations with less than $m + s$ parameters. Consequently, the RTS along the frontier may be very sensitive to (even small) changes in the data. Thus, for example, we may encounter cases were if we employ $1 + \delta$ for outputs $r = 1, \ldots, s - 1$, and $1 + 2 \cdot \delta$ for the last output we get a different RTS estimate than the one obtained by using $1 + \delta$ for all the outputs. This sensitivity can be addressed by qualifying the RTS outcome by some measure of the 'radius' around DMU_0 for which the RTS estimate still holds (see, e.g., the sensitivity models developed by Charnes et al., 1992, to qualify the efficiency rating of DMUs by the radius of the multi-dimensional ball around them in which they would still be considered efficient).

A related issue to the sensitivity of the RTS classifications is the question of determining the RTS magnitudes. The two work in opposite directions. The further the DMU is from a constant RTS

Table 2
Solutions of (7) for projections of inefficient DMUs ($\delta = 0.0001$)

DMU	X_1	X_2	Y	β_0^* (7)	RTS on the right
Input-oriented BCC model:					
2	1.1	1.1	1	1.00008182	IRS
4	1.91	1.91	1.9	1.00008953	IRS
9	4.125	4.575	3.9	1.00014918	DRS
Output-oriented BCC model:					
2	1.3	1.3	1.22222	1.00008446	IRS
4	2.1	2.1	2.1	1.0001	CRS
9	4.43	5	4.142857	1.000145	DRS

area of the frontier (in either the increasing or the decreasing RTS regions) the more insensitive is its RTS classification but the smaller is its scale efficiency. For practical purposes it is enough to observe the ratios of β/α for the efficient DMUs to rank them in order of their scale efficiency.

Once the RTS associated with a given unit are estimated one must ask how can this information be used to improve the operations of the unit. Naturally, one would like to consider expanding the operations of efficient DMUs operating under increasing RTS or contract the level of activity at DMUs operating under decreasing RTS. In other words, one would like to bring about movements of inefficient units onto the frontier and movements of efficient units along the frontier in desirable directions. We point out this topic as an important area for further research.

Appendix

The modified RTS estimation technique of Banker and Thrall (1992, p.81) solves the following two programs for each BCC efficient DMU_0:

$$\text{Max} \quad u_0 = u_0^+ \tag{A.1}$$

s.t.

$$U \cdot Y_0 + u_0 = 1,$$

$$U \cdot Y - V \cdot X + u_0 \leqslant 0,$$

$$V \cdot X_0 = 1,$$

$$U, V \geqslant 0,$$

$$\text{Min} \quad u_0 = u_0^- \tag{A.2}$$

s.t.

$$U \cdot Y_0 + u_0 = 1,$$

$$U \cdot Y - V \cdot X + u_0 \leqslant 0,$$

$$V \cdot X_0 = 1,$$

$$U, V \geqslant 0.$$

BT distinguish between 3 cases: 1) $u_0^- > 0$: increasing RTS; 2) $u_0^+ < 0$: decreasing RTS; 3) $u_0^- < 0$ and $u_0^+ > 0$: constant RTS.

Consider the following example with 2 DMUs, one input and one output (x, y): DMU_1 (0.5, 1), DMU_2 (2, 2), as shown in Fig. 7.

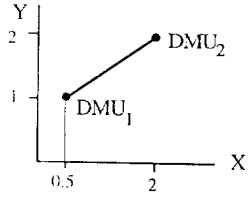

Fig. 7. An illustration of the Banker–Thrall technique.

Both DMUs are BCC efficient. Clearly, DRS prevail along the frontier which in this case consists only of the line segment connecting the two points. However, because of Postulate 2 of Banker and Thrall (1992, p.78) the line connecting the first DMU with the point (0.5, 0) is considered part of the frontier. Hence, when (A.1) and (A.2) are solved for DMU_1 we get $u_0^+ = 1$ (with $V = 2$ and $U = 0$) and $u_0^- = -2$ (with $V = 2$ and $U = 3$). Thus, according to Banker and Thrall, CRS prevail for DMU_1. In contrast, the method proposed here would clearly state that there is no information to determine the RTS (left and) below DMU_1 and that DRS prevail to the right (and above) of DMU_1. Similar situation would have occurred on the right hand side of the frontier if the values of DMU_1 were (1.5, 1). Again, the modified technique would declare DMU_2 as having CRS while it can really be associated here only with IRS.

References

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 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), "Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis", *Management Science* 30/9, 1078–1092.

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

Brinkerhoff, R.O., and Dressler, D.E. (1990), "Productivity measurement: A guide for managers and evaluators", in: *Applied Social Research Methods Series, Vol. 19*, Sage, London.

Chang, K.P., and Guh, Y.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 efficiency of decision making units", *European Journal of Operational Research* 2/6, 429–444.

Charnes, A., Haag, S., Jaska, P., and Semple, J. (1992), "Sensitivity of efficiency classifications in the additive model of Data Envelopment Analysis", *International Journal of System Sciences* 23/5, 789–798.

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

Färe, R., Grosskopf, S., and Lovell, C.A.K. (1985), *The measurement of Efficiency of Production*, Kluwer–Nijhoff, The Hague.

Seiford, L.M. (1991), "A bibliography of Data Envelopment Analysis" (1978–1991), Working Paper, The University of Massachusetts, Amherst, MA.