

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227220044>

# Returns to scale in DEA

Chapter · August 2011

DOI: 10.1007/978-1-4419-6151-8\_2

CITATIONS

67

READS

6,886

4 authors, including:



Lawrence M. Seiford  
University of Michigan

141 PUBLICATIONS 32,247 CITATIONS

[SEE PROFILE](#)



Joe Zhu  
Worcester Polytechnic Institute

279 PUBLICATIONS 25,689 CITATIONS

[SEE PROFILE](#)

## Chapter 2

# Returns to Scale in DEA\*

Rajiv D. Banker, William W. Cooper, Lawrence M. Seiford, and Joe Zhu

**Abstract** This chapter discusses returns to scale (RTS) in data envelopment analysis (DEA). The BCC and CCR models described in Chap. 1 of this handbook are treated in input-oriented forms, while the multiplicative model is treated in output-oriented form. (This distinction is not pertinent for the additive model, which simultaneously maximizes outputs and minimizes inputs in the sense of a vector optimization.) Quantitative estimates in the form of scale elasticities are treated in the context of multiplicative models, but the bulk of the discussion is confined to qualitative characterizations such as whether RTS is identified as increasing, decreasing, or constant. This is discussed for each type of model, and relations between the results for the different models are established. The opening section describes and delimits approaches to be examined. The concluding section outlines further opportunities for research and an Appendix discusses other approaches in DEA treatment of RTS.

**Keywords** Data envelopment analysis • Efficiency • Returns to scale

## 2.1 Introduction

It has long been recognized that Data Envelopment Analysis (DEA) by its use of mathematical programming is particularly adept at estimating inefficiencies in multiple input and multiple output production correspondences. Following Charnes, Cooper, and Rhodes (CCR 1978), a number of different DEA models

---

\*Part of the material in this chapter is adapted from European Journal of Operational Research, Vol 154, Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M. and Zhu, J, Returns to scale in different DEA models, 345–362, 2004, with permission from Elsevier Science.

J. Zhu (✉)

School of Business, Worcester Polytechnic Institute, Worcester, MA 01609, USA  
e-mail: [jzhu@wpi.edu](mailto:jzhu@wpi.edu)

have now appeared in the literature (see Cooper et al. 2000). During this period of model development, the economic concept of returns to scale (RTS) has also been widely studied within the different frameworks provided by these methods, and this is the topic to which this chapter is devoted.

In the literature of classical economics, RTS have typically been defined only for single-output situations. RTS are considered to be increasing if a proportional increase in all the inputs results in a more than proportional increase in the single output. Let  $\alpha$  represent the proportional input increase and  $\beta$  represent the resulting proportional increase of the single output. Increasing returns to scale (IRS) prevail if  $\beta > \alpha$ , and decreasing returns to scale (DRS) prevail if  $\beta < \alpha$ . Banker (1984), Banker et al. (1984), and Banker and Thrall (1992) extend the RTS concept from the single-output case to multiple-output cases using DEA.

Two paths may be followed in treating RTS in DEA. The first path, developed by Färe, Grosskopf, and Lovell (FGL 1985, 1994), determines RTS by a use of ratios of radial measures. These ratios are developed from model pairs which differ only in whether conditions of convexity and subconvexity are satisfied. The second path stems from work by Banker (1984), Banker et al. (1984) and Banker and Thrall (1992). This path, which is the one we follow, includes, but is not restricted to, radial measure models. It extends to additive and multiplicative models as well, and does so in ways that provide opportunities for added insight into the nature of RTS and its treatment by the methods and concepts of DEA.

The FGL approach has now achieved a considerable degree of uniformity that has long been available – as in FGL (1985), for instance. See also FGL (1994). We therefore treat their approach in the Appendix to this chapter. This allows us to center this chapter on treating RTS with different models. These treatments have long been available but only in widely scattered literatures. We also delineate relations that have been established between these different treatments and extend this to relations that have also been established with the FGL approach. See Banker et al. (1996b), Zhu and Shen (1995), and Färe and Grosskopf (1994). In particular, Seiford and Zhu (1999) established the relations among these alternative approaches and provided a simple approach to RTS estimation without the need for checking multiple optimal solutions.

The plan of development in this chapter starts with a recapitulation of results from the very important paper by Banker and Thrall (1992). Although developed in the context of radial measure models, we also use the Banker and Thrall (1992) results to unify the treatment of all of the models we cover. This is done after we first cover the radial measure models that are treated by Banker and Thrall (1992). Proofs of their theorems are not supplied because these are already available in Banker and Thrall (1992). Instead refinements from Banker et al. (1996a) and from Banker et al. (1996b) are introduced, which are directed to (a) providing simpler forms for implementing the Banker–Thrall theorems and (b) eliminating some of the assumptions underlying these theorems.

We then turn to concepts such as the MPSS (Most Productive Scale Size) introduced by Banker (1984) to treat multiple output–multiple input cases

in DEA and extend RTS concepts built around the single-output case in classical economics. Additive and multiplicative models are then examined, and the latter are used to introduce (and prove) new theorems for determining scale elasticities.

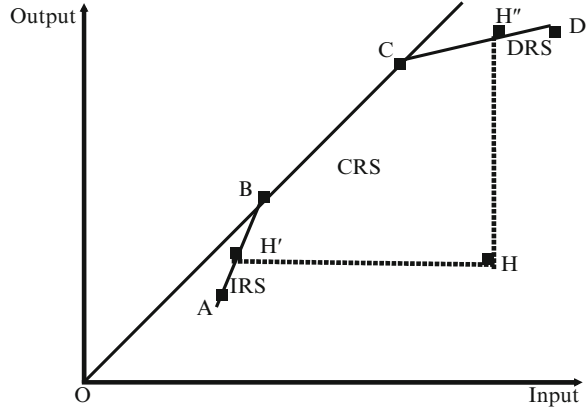
The former (i.e., the additive case) is joined with a “goal vector” approach introduced by Thrall (1996a) to make contact with “invariance” and “balance” ideas that play prominent roles in the “dimensional analysis” used to guide the measurements used in the natural sciences (such as physics). We next turn to the class of multiplicative models where, as shown by Charnes et al. (1982, 1983) and Banker and Maindiratta (1986), the piecewise linear frontiers usually employed in DEA are replaced by a frontier that is piecewise Cobb–Douglas (= log linear). Scale elasticity estimates are then obtained from the exponents of these “Cobb–Douglas like” functions for the different segments that form a frontier, which need not be concave. A concluding section points up issues for further research.

The [Appendix](#) of this chapter presents the FGL approach. We then present a simple RTS approach developed by Zhu and Shen (1995) and Seiford and Zhu (1999) to avoid the need for checking the multiple optimal solutions. This approach will substantially reduce the computational burden because it relies on the standard CCR and BCC computational codes.

## 2.2 RTS Approaches with BCC Models

For ease of reference, we present here the BCC models. Suppose that we have  $n$  DMUs (decision-making units) where every  $DMU_j, j = 1, 2, \dots, n$ , produces the same  $s$  outputs in (possibly) different amounts,  $y_{rj}$  ( $r = 1, 2, \dots, s$ ), using the same  $m$  inputs,  $x_{ij}$  ( $i = 1, 2, \dots, m$ ), also in (possibly) different amounts. The efficiency of a specific  $DMU_o$  can be evaluated by the “BCC model” of DEA in “envelopment form” as follows,

$$\begin{aligned}
 & \min \theta_o - \varepsilon \left( \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right), \\
 & \text{subject to} \\
 & \theta_o x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^- \quad i = 1, 2, \dots, m, \\
 & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ \quad r = 1, 2, \dots, s, \\
 & 1 = \sum_{j=1}^n \lambda_j, \\
 & 0 \leq \lambda_j, s_i^-, s_r^+ \quad \forall i, r, j,
 \end{aligned} \tag{2.1}$$

**Fig. 2.1** Returns to scale

where, as discussed for the expression (1.6) in Chap. 1,  $\varepsilon > 0$  is a non-Archimedean element defined to be smaller than any positive real number.

As noted in the abstract, we confine attention to input-oriented versions of these radial measure models and delay discussion of changes in mix, as distinct from changes in scale, until we come to the class of additive models where input and output orientations are treated simultaneously. Finally, we do use output orientations in the case of multiplicative models because, as will be seen, the formulations in that case do not create problems in distinguishing between scale and mix changes.

*Remark.* We should, however, note that input- and output-oriented models may give different results in their RTS findings. See Fig. 2.1 and related discussion below. Thus the result secured may depend on the orientation used. IRS may result from an input-oriented model, for example, while an application of an output-oriented model may produce a DRS characterization from the same data. See Golany and Yu (1994) for treatments of this problem.

The dual (multiplier) form of the BCC model represented in (2.1) is obtained from the same data that are then used in the following form:

$$\begin{aligned}
 \max z &= \sum_{r=1}^s u_r y_{ro} - u_o, \\
 \text{subject to} \\
 \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - u_o &\leq 0, \quad j = 1, \dots, n, \\
 \sum_{i=1}^m v_i x_{io} &= 1, \\
 v_i &\geq \varepsilon, \quad u_r \geq \varepsilon, \quad u_o \text{ free in sign.}
 \end{aligned} \tag{2.2}$$

The above formulations assume that  $x_{ij}, y_{rj} \geq 0 \quad \forall i, r, j$ . All variables in (2.2) are also constrained to be nonnegative – except for  $u_o$ , which may be positive, negative, or zero with consequences that make it possible to use optimal values of this variable to identify RTS.

When a  $DMU_o$  is efficient in accordance with the Definition 1.3 in Chap. 1, the optimal value of  $u_o$ , i.e.,  $u_o^*$ , in (2.2), can be used to characterize the situation for RTS.

RTS generally has an unambiguous meaning only if  $DMU_o$  is on the efficiency frontier – since it is only in this state that a tradeoff between inputs and outputs is *required* to improve one or the other of these elements. However, there is no need to be concerned about the efficiency status in our analyses because efficiency can always be achieved as follows. If a  $DMU_o$  is not BCC efficient, we can use optimal values from (2.1) to project this DMU onto the BCC efficiency frontier via the following formulas,

$$\begin{cases} \hat{x}_{io} = \theta_o^* x_{io} - s_i^{-*} = \sum_{j=1}^n x_{ij} \lambda_j^*, & i = 1, \dots, m, \\ \hat{y}_{ro} = y_{ro} + s_r^{+*} = \sum_{j=1}^n y_{rj} \lambda_j^*, & r = 1, \dots, s, \end{cases} \quad (2.3)$$

where the symbol “\*” denotes an optimal value. These are sometimes referred to as the “CCR Projection Formulas” because Charnes et al. (1978) showed that the resulting  $\hat{x}_{io} \leq x_{io}$  and  $\hat{y}_{ro} \geq y_{ro}$  correspond to the coordinates of a point on the efficiency frontier. They are, in fact, coordinates of the point used to evaluate  $DMU_o$  when (2.1) is employed.

Suppose that we have five DMUs,  $A, B, C, D$ , and  $H$  as shown in Fig. 2.1 from Zhu (2009). Ray  $OBC$  is the constant returns to scale (CRS) frontier.  $AB, BC$ , and  $CD$  constitute the BCC frontier, and exhibit increasing, constant, and decreasing returns to scale, respectively.  $B$  and  $C$  exhibit CRS. On the line segment  $AB$ , IRS prevail to the left of  $B$  for the BCC model and on the line segment  $CD$ , DRS prevail to the right of  $C$ . By applying (2.3) to point  $H$ , we have a frontier point  $H'$  on the line segment  $AB$  where IRS prevail. However, if we use the output-oriented BCC model, the projection is onto  $H''$  where DRS prevail. This is due to the fact that the input-oriented and the output-oriented BCC models yield different projection points on the BCC frontier and it is on the frontier that RTS is determined. See Zhu (2009) for discussion on “returns to scale regions.”

We now present our theorem for RTS as obtained from Banker and Thrall (1992, p. 79) who identify RTS with the sign of  $u_o^*$  in (2.2) as follows:

**Theorem 2.1.** The following conditions identify the situation for RTS for the BCC model given in (2.2),

- (i) IRS prevail at  $(\hat{x}_o, \hat{y}_o)$  if and only if  $u_o^* < 0$  for all optimal solutions.
- (ii) DRS prevail at  $(\hat{x}_o, \hat{y}_o)$  if and only if  $u_o^* > 0$  for all optimal solutions.
- (iii) Constant RTS prevail at  $(\hat{x}_o, \hat{y}_o)$  if and only if  $u_o^* = 0$  for at least one optimal solution.

Here, it may be noted,  $(\hat{x}_o, \hat{y}_o)$  are the coordinates of the point on the efficiency frontier which is obtained from (2.3) in the evaluation of  $DMU_o$  via the solution to (2.1). Note, therefore, that a use of the projection makes it unnecessary to *assume* that the points to be analyzed are all on the BCC efficient frontier – as was assumed in Banker and Thrall (1992).

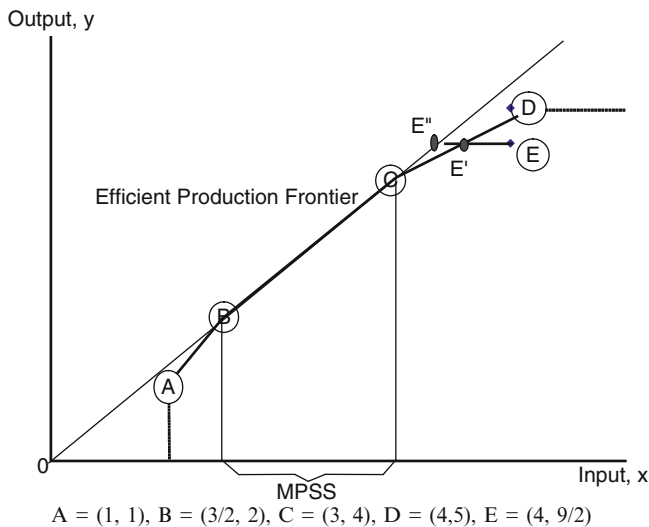
An examination of all optimal solutions can be onerous. Therefore, Banker and Thrall (1992) provide one way of avoiding a need for examining *all* optimal solutions. However, Banker et al. (1996a) approach is used here because it avoids the possibility of infinite solutions that are present in the Banker–Thrall approach. In addition, the Banker et al. (1996a) approach insures that the RTS analyses are conducted on the efficiency frontier. This is accomplished as follows.

Suppose that an optimum has been achieved with  $u_o^* < 0$ . As suggested by Banker et al. (1996a), the following model may then be employed to avoid having to explore all alternate optima,

$$\begin{aligned}
 & \text{maximize } \hat{u}_o, \\
 & \text{subject to} \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - \hat{u}_o \leq 0, \quad j = 1, \dots, n; j \neq 0, \\
 & \sum_{r=1}^s u_r \hat{y}_{ro} - \sum_{i=1}^m v_i \hat{x}_{io} - \hat{u}_o \leq 0, \quad j = 0, \\
 & \sum_{i=1}^m v_i \hat{x}_{io} = 1, \\
 & \sum_{r=1}^s u_r \hat{y}_{ro} - \hat{u}_o = 1, \\
 & v_i, u_r \geq 0 \quad \text{and} \quad \hat{u}_o \leq 0,
 \end{aligned} \tag{2.4}$$

where the  $\hat{x}_{io}$  and  $\hat{y}_{ro}$  are obtained from (2.3).

With these changes of data, the constraints for (2.4) are in the same form as (2.2) except for the added conditions  $\sum_{r=1}^s u_r \hat{y}_{ro} - \hat{u}_o = 1$  and  $\hat{u}_o \leq 0$ . The first of these conditions helps to ensure that we will be confined to the efficiency frontier. The second condition allows us to determine whether an optimal value can be achieved with  $\max \hat{u}_o = 0$ . If  $\hat{u}_o^* = 0$  can be obtained, then condition (iii) of Theorem 2.1 is satisfied and RTS are constant. If, however,  $\max \hat{u}_o = \hat{u}_o^* < 0$ , then, as set forth in (i) of Theorem 2.1, RTS are increasing. In either case, the problem is resolved, and the need for examining all alternate optima is avoided in this way of implementing Theorem 2.1.



**Fig. 2.2** Most productive scale size

We can deal in a similar manner with the case when  $u_o^* > 0$  by (a) reorienting the objective in (2.4) to “minimize”  $\hat{u}_o$  and (b) replacing the constraint  $\hat{u}_o \leq 0$  with  $\hat{u}_o \geq 0$ . All other elements of (2.4) remain the same and if  $\min \hat{u}_o = \hat{u}_o^* > 0$  then condition (ii) of Theorem 2.1 is applicable while if  $\min \hat{u}_o = \hat{u}_o^* = 0$  then condition (iii) is applicable.

Reference to Fig. 2.2 can help us to interpret these results. This figure portrays the case of one input,  $x$ , and one output,  $y$ . The coordinates of each point are listed in the order  $(x, y)$ . Now, consider the data for A that has the coordinates  $(x = 1, y = 1)$ , as shown at the bottom of Fig. 2.2. The “supports” at A form a family that starts at the vertical line (indicated by the dotted line) and continues through rotations about A until coincidence is achieved with the line connecting A to B. All of these supports will have negative intercepts so  $u_o^* < 0$  and the situation for A is one of IRS as stated in (i) of Theorem 2.1.

The reverse situation applies at D. Starting with the horizontal line indicated by the dots, supports can be rotated around D until coincidence is achieved with the line connecting D and C. In all cases, the intercept is positive so  $u_o^* > 0$  and RTS are decreasing as stated in (ii) of Theorem 2.1.

Rotations at C or B involve a family of supports in which at least one member will achieve coincidence with the broken line going through the origin so that, in at least this single case, we will have  $u_o^* = 0$ , in conformance with the condition for constant RTS in (iii) of Theorem 2.1.

Finally, we turn to E, the only point that is BCC inefficient in Fig. 2.2. Application of (2.3), however, projects E into  $E'$  – a point on the line between C and D – and, therefore, gives the case of DRS with a unique solution of  $\hat{u}_o^* > 0$ .



Hence, all possibilities are comprehended by Theorem 2.1 for the qualitative RTS characterizations that are of concern here. Thus, for these characterizations only the signs of the nonzero values of  $\hat{u}_o^*$  suffice.

## 2.3 RTS Approaches with CCR Models

We now turn to the CCR models that, as discussed in Chap. 1 of this handbook, take the following form,

$$\begin{aligned}
 & \text{minimize } \theta - \varepsilon \left( \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right), \\
 & \text{subject to} \\
 & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \\
 & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \\
 & 0 \leq \lambda_j, s_i^-, s_r^+ \quad \forall i, j, r.
 \end{aligned} \tag{2.5}$$

As can be seen, this model is the same as the “envelopment form” of the BCC model in (2.1) except for the fact that the condition  $\sum_{j=1}^n \lambda_j = 1$  is omitted. In consequence, the variable  $u_o$ , which appears in the “multiplier form” for the BCC model in (2.2), is omitted from the dual (multiplier) form of this CCR model. The projection formulas expressed in (2.3) are the same for both models. We can, therefore, use these same projections to move all points onto the efficient frontier for (2.5) and proceed directly to RTS characterizations for (2.5), which are supplied by the following theorem from Banker and Thrall (1992).

**Theorem 2.2.** The following conditions identify the situation for RTS for the CCR model given in (2.5):

- (i) CRS prevail at  $(\hat{x}_o, \hat{y}_o)$  if  $\sum \lambda_j^* = 1$  in any alternate optimum.
- (ii) DRS prevail at  $(\hat{x}_o, \hat{y}_o)$  if  $\sum \lambda_j^* > 1$  for all alternate optima.
- (iii) IRS prevail at  $(\hat{x}_o, \hat{y}_o)$  if  $\sum \lambda_j^* < 1$  for all alternate optima.

Following Banker et al. (1996b), we can avoid the need for examining all alternate optima. This is done as follows. Suppose that an optimum has been obtained for (2.5) with  $\sum \lambda_j^* < 1$ . We then replace (2.5) with

$$\begin{aligned}
& \text{maximize } \sum_{j=1}^n \hat{\lambda}_j + \varepsilon \left( \sum_{i=1}^m \hat{s}_i^- + \sum_{r=1}^s \hat{s}_r^+ \right), \\
& \text{subject to} \\
& \theta^* x_{io} = \sum_{j=1}^n x_{ij} \hat{\lambda}_j + \hat{s}_i^- \quad \text{for } i = 1, \dots, m, \\
& y_{ro} = \sum_{j=1}^n y_{rj} \hat{\lambda}_j - \hat{s}_r^+ \quad \text{for } r = 1, \dots, s, \\
& 1 \geq \sum_{j=1}^n \hat{\lambda}_j, \\
& \text{with } 0 \leq \hat{\lambda}_j, \hat{s}_i^-, \hat{s}_r^+ \quad \forall i, j, r,
\end{aligned} \tag{2.6}$$

where  $\theta^*$  is the optimal value of  $\theta$  secured from (2.5).

*Remark.* This model can also be used for setting scale-efficient targets when multiple optimal solutions in model (2.5) are present. See Zhu (2000, 2002).

We note for (2.5) that we may omit the two-stage process described for the CCR model in Chap. 1 – i.e., the process in which the sum of the slacks are maximized in stage 2 after  $\theta^*$  has been determined. This is replaced with a similar two-stage process for (2.6) because only the optimal value of  $\theta$  is needed from (2.5) to implement the analysis now being described. The optimal solution to (2.6) then yields values of  $\hat{\lambda}_j^*$ ,  $j = 1, \dots, n$ , for which the following theorem is immediate,

**Theorem 2.3.** Given the existence of an optimal solution with  $\sum \lambda_j^* < 1$  in (2.5), the RTS at  $(\hat{x}_o, \hat{y}_o)$  are constant if and only if  $\sum \lambda_j^* = 1$  and RTS are increasing if and only if  $\sum \lambda_j^* < 1$  in (2.6).

Consider  $A = (1, 1)$  as shown at the bottom of Fig. 2.2. Because we are only interested in  $\theta^*$ , we apply (1.4) in Chap. 1 to obtain

$$\begin{aligned}
& \text{minimize } \theta, \\
& \text{subject to} \\
& 1\theta \geq 1\lambda_A + \frac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E, \\
& 1 \leq 1\lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \frac{9}{2}\lambda_E, \\
& 0 \leq \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E.
\end{aligned} \tag{2.7}$$

This problem has  $\theta^* = 3/4$  and hence  $A$  is found to be inefficient. Next, we observe that this problem has alternate optima because this same  $\theta^* = 3/4$  can be obtained from either  $\lambda_B^* = 1/2$  or from  $\lambda_C^* = 1/4$  with all other  $\lambda^* = 0$ . For each of these optima, we have  $\sum \lambda_j^* < 1$ , so we utilize (2.6) and write

$$\begin{aligned}
& \text{maximize } \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E + \varepsilon(\hat{s}^- + s^+), \\
& \text{subject to} \\
& \frac{3}{4} = 1\hat{\lambda}_A + \frac{3}{2}\hat{\lambda}_B + 3\hat{\lambda}_C + 4\hat{\lambda}_D + 4\hat{\lambda}_E + \hat{s}^-, \\
& 1 = 1\hat{\lambda}_A + 2\hat{\lambda}_B + 4\hat{\lambda}_C + 5\hat{\lambda}_D + \frac{9}{2}\hat{\lambda}_E - \hat{s}^+, \\
& 1 \geq \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E, \\
& 0 \leq \hat{\lambda}_A, \hat{\lambda}_B, \hat{\lambda}_C, \hat{\lambda}_D, \hat{\lambda}_E.
\end{aligned} \tag{2.8}$$

so that  $\sum \hat{\lambda}_j^* \equiv \hat{\lambda}_A^* + \hat{\lambda}_B^* + \hat{\lambda}_C^* + \hat{\lambda}_D^* + \hat{\lambda}_E^*$  with all  $\hat{\lambda}$  nonnegative. An optimal solution is  $\hat{\lambda}_B^* = 1/2$  and all other  $\hat{\lambda}^* = 0$ . Hence,  $\sum \hat{\lambda}_j^* < 1$ , so from Theorem 2.3, IRS prevail at A.

We are here restricting attention to solutions of (2.6) with  $\sum_{j=1}^n \hat{\lambda}_j \leq 1$ , as in the constraint of (2.6), but the examples we provide below show how to treat situations in which  $\theta^*$  is associated with solutions of (2.5) that have values  $\sum \hat{\lambda}_j^* > 1$ .

Consider  $E = (4, 9/2)$  for (2.6), as a point that is not on either (i) the BCC efficiency frontier represented by the solid lines in Fig. 2.2 or (ii) the CCR efficiency frontier represented by the broken line from the origin. Hence, both the BCC and CCR models find  $E$  to be inefficient. Proceeding via the CCR envelopment model in (2.5) with the slacks omitted from the objective, we get

$$\begin{aligned}
& \text{minimize } \theta, \\
& \text{subject to} \\
& 4\theta \geq 1\lambda_A + \frac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E, \\
& \frac{9}{2} \leq 1\lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \frac{9}{2}\lambda_E, \\
& 0 \leq \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E.
\end{aligned} \tag{2.9}$$

Again we have alternate optima with, now,  $\theta^* = 27/32$  for either  $\lambda_B^* = 9/4$  or  $\lambda_C^* = 9/8$  and all other  $\lambda^* = 0$ . Hence, in both cases, we have  $\sum \hat{\lambda}_j^* > 1$ . Continuing in an obvious way, we next reorient the last constraint and the objective in (2.6) to obtain

$$\begin{aligned}
& \text{minimize } (\hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E) - \varepsilon(\hat{s}^- + s^+), \\
& \text{subject to} \\
& \frac{27}{8} = 1\hat{\lambda}_A + \frac{3}{2}\hat{\lambda}_B + 3\hat{\lambda}_C + 4\hat{\lambda}_D + 4\hat{\lambda}_E + \hat{s}^-, \\
& \frac{9}{2} = 1\hat{\lambda}_A + 2\hat{\lambda}_B + 4\hat{\lambda}_C + 5\hat{\lambda}_D + \frac{9}{2}\hat{\lambda}_E - \hat{s}^+, \\
& 1 \leq \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E, \\
& 0 \leq \hat{\lambda}_A, \hat{\lambda}_B, \hat{\lambda}_C, \hat{\lambda}_D, \hat{\lambda}_E.
\end{aligned} \tag{2.10}$$

This has its optimum at  $\hat{\lambda}_C^* = 9/8$  with all other  $\hat{\lambda}^* = 0$ . So, in conformance with Theorem 2.3, as given for (2.6), we associate  $E$  with DRS.

There is confusion in the literature on the RTS characterizations obtained from Theorems 2.1 and 2.2 and the BCC and the CCR models with which they are associated. Hence, we proceed further as follows.

As noted earlier, RTS generally has an unambiguous meaning only for points on the efficiency frontier. When the BCC model as given in (2.1) is used on the data in Fig. 2.2, the primal model projects  $E$  into  $E'$  with coordinates  $(7/2, 9/2)$  on the segment of the line  $y = 1 + x$ , which connects  $C$  to  $D$  on the BCC efficiency frontier. Comparing this with  $E = (4, 9/2)$  identifies  $E$  as having an inefficiency in the amount of  $1/2$  unit in its input. This is a technical inefficiency, in the terminology of DEA. Turning to the dual for  $E$  formed from the BCC model, as given in (2.2), we obtain  $u_o^* = 1/4$ . Via Theorem 2.1 this positive value of  $u_o^*$  suggests that RTS are either decreasing or constant at  $E' = (28/8, 9/2)$  – the point to which  $E$  is projected to obtain access to model (2.4). Substitution in the latter model yields a value of  $\hat{u}_o^* \doteq 2/7$ , which is also positive, thereby identifying  $E'$  with the DRS that prevail for the BCC model on this portion of the efficiency frontier in Fig. 2.2.

Next, we turn to the conditions specified in Theorem 2.2, which are identified with the CCR envelopment model (2.5). Here, we find that the projection is to a new point  $E'' = (27/8, 9/2)$ , which is on the line  $y = 4/3x$  corresponding to the broken line from the origin that coincides with the segment from  $B$  to  $C$  in Fig. 2.2. This ray from the origin constitutes the efficiency frontier for the CCR model, which, when used in the manner we have previously indicated, simultaneously evaluates the technical and RTS performances of  $E$ . In fact, as can be seen from the solution to (2.9), this evaluation is effected by either  $\hat{\lambda}_B^* = 9/4$  or  $\hat{\lambda}_C^* = 9/8$  – which are variables associated with vectors in a “CRS region” that we will shortly associate with “most productive scale size” (MPSS) for the BCC model. The additional  $1/8$  unit input reduction effected in going from  $E''$  to  $E'$  is needed to adjust to the efficient mix that prevails in this MPSS region which the CCR model is using to evaluate  $E$ .

Thus, the CCR model as given in (2.5) simultaneously evaluates scale and purely technical inefficiencies, while the BCC model, as given in (2.1), separates out the scale inefficiencies for evaluation in its associated dual (=multiplier) form as given in (2.2). Finally, as is well known, a simplex method solution to (2.1) automatically supplies the solution to its dual in (2.2). Thus, no additional computations are required to separate the purely technical inefficiency characterizations obtained from (2.1) and the RTS characterizations obtained from (2.2). Both sets of values are obtainable from a solution to (2.1).

We now introduce the following theorem that allows us to consider the relations between Theorems 2.1 and 2.2 in the RTS characterization.

**Theorem 2.4.** Suppose that  $DMU_o$  is designated as efficient by the CCR model,  $DMU_o$  then it is also designated as efficient by the BCC model.

*Proof.* The CCR and BCC models differ only because the latter has the additional constraint  $\sum_{j=1}^n \lambda_j = 1$ . The following relation must therefore hold

$$\theta_{\text{CCR}}^* - \varepsilon \left( \sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right) \leq \theta_{\text{BCC}}^* - \varepsilon \left( \sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right),$$

where the expressions on the left and right of the inequality respectively designate optimal values for objective of the CCR and BCC models.

Now, suppose DMU<sub>o</sub> is found to be efficient with the CCR model. This implies  $\theta_{\text{CCR}}^* = 1$  and all slacks are zero for the expression on the left. Hence, we will have

$$1 \leq \theta_{\text{BCC}}^* - \varepsilon \left( \sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right).$$

However, the  $x_{io}$  and  $y_{ro}$  values appear on both the left and right sides of the corresponding constraints in the DEA models. Hence, choosing  $\lambda_o^* = \lambda_j^* = 1$ , we can always achieve equality with  $\theta_{\text{BCC}}^*$  which is the lower bound in this exact expression with all slacks zero. Thus, this DMU<sub>o</sub> will also be characterized as efficient by the BCC model whenever it is designated as efficient by the CCR model. ■

We now note that the reverse of this theorem is not true. That is, a DMU<sub>o</sub> may be designated as efficient by the BCC model but not by the CCR model. Even when both models designate a DMU<sub>o</sub> as inefficient, moreover, the measures of inefficiency may differ. Application of the BCC model to point *E* in Fig. 2.2, for example, will designate *E''* on the line connecting *C* and *D* to evaluate its efficiency. However, utilization of the CCR model will designate *E'* with  $\theta_{\text{CCR}}^* < \theta_{\text{BCC}}^*$  so that  $1 - \theta_{\text{CCR}}^* > 1 - \theta_{\text{BCC}}^*$ , which shows a greater inefficiency value for DMU<sub>o</sub> when the CCR model is used.

Because DEA evaluates relative efficiency, it will always be the case that at least one DMU will be characterized as efficient by either model. However, there will always be at least one point of intersection between these two frontiers. Moreover, the region of the intersection will generally expand a DMU set to be efficient with the CCR model. The greatest spread between the envelopments will then constitute extreme points that define the boundaries of the intersection between the CCR and BCC models.

The way Theorem 2.2 effects its efficiency characterization is by models of linear programming algorithms that use “extreme point” methods. That is, the solutions are expressed in terms of “basis sets” consisting of extreme points. The extreme points *B* and *C* in Fig. 2.2 can constitute active members of such an “optimal basis” where, by “active member,” we refer to members of a basis that have nonzero coefficients in an optimal solution.

Now, as shown in Cooper et al. (2000), active members of an optimal basis are necessarily efficient. For instance,  $\lambda_B^* = 1/2$  in the solution to (2.8) designates *B* as

an active member of an optimal basis and the same is true for  $\lambda_C^* = 1/4$ , and both  $B$  and  $C$  are therefore efficient. In both cases, we have  $\sum \lambda_j^* < 1$  and we have IRS at point  $(3/4, 1)$  on the CRS ray, which is used to evaluate  $A$ . In other words,  $\sum \lambda_j^* < 1$  shows that  $B$  and  $C$  both lie below the region of intersection because its coordinates are smaller in value than the corresponding values of the active members in the optimal basis.

Turning to the evaluation of  $E$ , we have  $\theta^* = 27/32 < 1$  showing that  $E$  is inefficient in (2.10). Either  $\lambda_B^* = 9/4$  or  $\lambda_C^* = 9/8$  can serve as active member in the basis. Thus, to express  $E''$  in terms of these bases, we have

$$\underset{E''}{\left(\frac{27}{8}, \frac{9}{2}\right)} = \underset{B}{\frac{9}{4} \left(\frac{3}{2}, 2\right)} = \underset{C}{\frac{9}{8} (3, 4)}$$

because  $E''$  lies above the region of intersection, as shown by  $\sum \lambda_j^* > 1$  for either of the optimal solutions.

As shown in the next section of this chapter, Banker (1984) refers to the region between  $B$  and  $C$  as the region of most productive scale size (MPSS). For justification, we might note that the slope of the ray from the origin through  $B$  and  $C$  is steeper than the slope of any other ray from the origin that intersects the production possibility set (PPS). This measure with the output per unit input is maximal relative to any other ray that intersects PPS.

Hence, Theorem 2.2 is using the values of  $\sum \lambda_j^*$  to determine whether RTS efficiency has achieved MPSS and what needs to be done to express this relative to the region of MPSS. We can, therefore, conclude with a corollary to Theorem 2.4:  $DMU_o$  will be at MPSS if and only if  $\sum \lambda_j^* = 1$  in an optimal solution when it is evaluated by a CCR model.

To see how this all comes about mathematically and how it relates to the RTS characterization, we note that the optimal solution for the CCR model consists of all points on the ray from the origin that intersect the MPSS region. If the point being evaluated is in MPSS, it can be expressed as a convex combination of the extreme points of MPSS so that  $\sum \lambda_j^* = 1$ . If the point is above the region, its coordinate values will all be larger than their corresponding coordinates in MPSS so that we will have  $\sum \lambda_j^* > 1$ . If the point is below the region, we will have  $\sum \lambda_j^* < 1$ . Because the efficient frontier, as defined by the BCC model, is strictly concave, the solution will designate this point as being in the region of constant, decreasing, or increasing RTS, respectively.

Thus, the CCR model simultaneously evaluates RTS and technical inefficiency while the BCC model separately evaluates technical efficiency with  $\theta_{BCC}^*$  from the envelopment model and RTS with  $u_o^*$  obtained from the multiplier model. As Fig. 2.2 illustrates, at point  $E$ , the evaluation for the CCR model is global with RTS always evaluated relative to MPSS. The evaluation for the BCC model is local with  $u_o^*$  being determined by the facet of the efficient frontier in which the

point used to evaluate  $DMU_o$  is located. As a consequence, it always be the case that  $\theta_{CCR}^* < \theta_{BCC}^*$  unless the point used to evaluate  $DMU_o$  is in the region of MPSS, in which case  $\theta_{CCR}^* = \theta_{BCC}^*$  will obtain.

## 2.4 Most Productive Scale Size

There is some ambiguity in dealing with points like  $B$  and  $C$  in Fig. 2.2 because the condition that prevails depends on the direction in which movement is to be effected. As noted by Førsund (1996) this situation was dealt with by Ragnar Frisch – who pioneered empirical studies of production and suggested that the orientation should be toward maximizing the output per unit input when dealing with technical conditions of efficiency. See Frisch (1964). However, Frisch (1964) dealt only with the case of single outputs. Extensions to multiple output–multiple input situations can be dealt with by the concept of Most Productive Scale Size (MPSS) as introduced into the DEA literature by Banker (1984). To see what this means consider

$$(X_o\alpha, Y_o\beta), \quad (2.11)$$

with  $\beta, \alpha \geq 0$  representing scalars and  $X_o$  and  $Y_o$  representing input and output vectors, respectively, for  $DMU_o$ . We can continue to move toward a possibly better (i.e., more productive) RTS situation as long as  $\beta/\alpha \neq 1$ . In other words, we are not at a point which is MPSS when either (a) all outputs can be increased in proportions that are at least as great as the corresponding proportional increases in all inputs needed to bring them about, or (b) all inputs can be decreased in proportions that are at least as great as the accompanying proportional reduction in all outputs. Only when  $\beta/\alpha = 1$ , or  $\alpha = \beta$ , will RTS be constant, as occurs at MPSS.

One way to resolve problems involving returns to scale for multiple output–multiple input situations would use a recourse to prices, costs (or similar weights) to determine a “best” or “most economical” scale size. Here, however, we are using the concept of MPSS in a way that avoids the need for additional information on unit prices, costs, etc., by allowing all inputs and outputs to vary simultaneously in the proportions prescribed by  $\alpha$  and  $\beta$  in (2.11). Hence, MPSS allows us to continue to confine attention to purely technical inefficiencies, as before, while allowing for other possible choices after scale changes and size possibilities have been identified and evaluated in our DEA analyses.

The interpretation we have just provided for (2.11) refers to RTS locally, as is customary – e.g., in economics. However, this does not exhaust the uses that can be made of Banker’s (1984) MPSS. For instance, we can now replace our preceding local interpretation of (2.11) by one which is oriented globally. That is, we seek to characterize the returns to scale conditions for  $DMU_o$  with respect to MPSS instead of restricting this evaluation to the neighborhood of the point  $(X_o, Y_o)$  where, say, a derivative is to be evaluated. See Varian (1984, p. 20) for economic interpretations of restrictions needed to justify uses of derivatives. We also do

this in a way that enables us to relate Theorems 2.1 and 2.2 to each other and thereby provide further insight into how the BCC and CCR models relate to each other in scale size (and other) evaluations.

For these purposes, we introduce the following formulation,

$$\begin{aligned}
 & \text{maximize } \frac{\beta}{\alpha}, \\
 & \text{subject to} \\
 & \beta Y_o \leq \sum_{j=1}^n Y_j \lambda_j, \\
 & \alpha X_o \geq \sum_{j=1}^n X_j \lambda_j, \\
 & 1 = \sum_{j=1}^n \lambda_j, \\
 & 0 \leq \beta, \alpha, \lambda_j, \quad j = 1, \dots, n.
 \end{aligned} \tag{2.12}$$

Now note that the condition  $\sum \lambda_j = 1$  appears just as it does in (2.1). However, in contrast to (2.1), we are now moving to a global interpretation by jointly maximizing the proportional increase in outputs and minimizing the proportional decrease in inputs. We are also altering the characterizations so that these  $\alpha$  and  $\beta$  values now yield new vectors  $\hat{X}_o = \alpha X_o$  and  $\hat{Y}_o = \beta Y_o$ , which we can associate with points which are MPSS, as in the following

**Theorem 2.5.** A necessary condition for  $DMU_o$ , with output and input vectors  $Y_o$  and  $X_o$ , to be MPSS is  $\max \beta/\alpha = 1$  in (2.12), in which case RTS will be constant.

Theorem 2.5 follows from the fact that  $\beta = \alpha = 1$  with  $\lambda_j = 0$ ,  $\lambda_o = 1$  for  $j = o$  is a solution of (2.12), so that, always,  $\max \beta/\alpha = \beta^*/\alpha^* \geq 1$ . See the appendix in Cooper et al. (1996) for a proof and a reduction of (2.12) to a linear programming equivalent.

We illustrate with  $D = (4, 5)$  in Fig. 2.2 for which we utilize (2.12) to obtain

$$\begin{aligned}
 & \text{Maximize } \frac{\beta}{\alpha}, \\
 & \text{subject to} \\
 & 5\beta \leq 1\lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \frac{9}{2}\lambda_E, \\
 & 4\alpha \geq 1\lambda_A + \frac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E, \\
 & 1 = \lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E, \\
 & 0 \leq \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E.
 \end{aligned} \tag{2.13}$$



This has an optimum at  $\lambda_B^* = 1$  with  $\alpha^* = 3/8$  and  $\beta^* = 2/5$  to give  $\beta^*/\alpha^* = 16/15 > 1$ . Thus, MPSS is not achieved. Substituting in (2.13) with  $\lambda_B^* = 1$ , we can use this solution to obtain  $4\alpha^* = 3/2$  and  $5\beta^* = 2$  which are the coordinates of  $B$  in Fig. 2.2. Thus,  $D = (4,5)$  is evaluated globally by reference to  $B = (3/2,2)$ , which is in the region of CRS and hence is MPSS.

There is also an alternate optimum to (2.13) with  $\lambda_C^* = 1$  and  $\alpha^* = 3/4$ ,  $\beta^* = 4/5$  so, again,  $\beta^*/\alpha^* = 16/15$ , and  $D$  is not at MPSS. Moreover,  $4\alpha^* = 5$ ,  $5\beta^* = 4$  gives the coordinates of  $C = (3, 4)$ . Thus,  $D$  is again evaluated globally by a point in the region of MPSS. Indeed, any point in this region of MPSS would give the same value of  $\beta^*/\alpha^* = 16/15$ , since all such points are representable as convex combinations of  $B$  and  $C$ .

**Theorem 2.6.** Sign conditions for BCC and CCR models:

- (i) The case of IRS.  $u_o^* < 0$  for all optimal solutions to (2.2) if and only if  $\left(\sum \lambda_j^* - 1\right) < 0$  for all optimal solutions to (2.5).
- (ii) The case of DRS.  $u_o^* > 0$  for all optimal solutions to (2.2) if and only if  $\left(\sum \lambda_j^* - 1\right) > 0$  for all optimal solutions to (2.5).
- (iii) The case of CRS.  $u_o^* = 0$  for some optimal solutions to (2.2) if and only if  $\left(\sum \lambda_j^* - 1\right) = 0$  for some optimal solution to (2.5).

This theorem removes the possibility that uses of the CCR and BCC models might lead to different RTS characterizations. It is also remarkable because differences might be expected from the fact that (2.2) effects its evaluations locally with respect to a neighboring facet while (2.5) effects its evaluations globally with respect to a facet (or point) representing MPSS.

To see what this means we focus on active members of an optimal solution set as follows.

Turning to  $E$  in Fig. 2.2 we see that it is evaluated by  $E'$  when (2.1) is used. This point, in turn, can be represented as a convex combination of  $C$  and  $D$  with both of the latter vectors constituting active members of the optimal basis. The associated support coincides with the line segment connecting  $C$  and  $D$  with a (unique) value  $u_o^* > 0$  so RTS are decreasing, as determined from (2.2). This is a local evaluation. When (2.5) is used, the projection is to  $E''$ , with alternate optima at  $B$  or  $C$  respectively serving as the only active member of the optimal basis. Hence the evaluation by the CCR model is effected globally. Nevertheless, the same DRS characterization is secured.

We now note that  $E''$  may be projected into the MPSS region by means of the following formulas,

$$\frac{\theta^* x_{io} - s_i^{-*}}{\sum_{j=1}^n \hat{\lambda}_j^*},$$

$$\frac{y_{ro} + s_i^{+*}}{\sum_{j=1}^n \hat{\lambda}_j^*}, \quad (2.14)$$

where the denominators are secured from (2.6). This convexification of (2.3), which is due to Banker and Morey (1986), provides a different projection than (2.3). We illustrate for  $E''$  by using the solutions for (2.9) to obtain

$$\frac{4\theta^* - s_i^{-*}}{9/4} = \frac{27/8}{9/4} = 3/2,$$

$$\frac{y_{ro} + s_i^{+*}}{9/4} = \frac{9/2}{9/4} = 2.$$

This gives the coordinates of  $B$  from one optimal solution. The other optimal solution yields the coordinates of  $C$  via

$$\frac{4\theta^* - s_i^{-*}}{9/8} = \frac{27/8}{9/8} = 3,$$

$$\frac{y_{ro} + s_i^{+*}}{9/8} = \frac{9/2}{9/8} = 4.$$

This additional step brings us into coincidence with the results already described for the MPSS model given in (2.13). Consistency is again achieved even though the two models proceed by different routes. The MPSS model in (2.12) bypasses the issue of increasing vs. decreasing RTS and focuses on the issue of MPSS, but this same result can be achieved for (2.5) by using the additional step provided by the projection formula (2.14).

## 2.5 Additive Models

The model (2.12), which we used for MPSS, avoids the problem of choosing between input and output orientations, but this is not the only type of model for which this is true. The additive models to be examined in this section also have this property. That is, these models simultaneously maximize outputs and minimize inputs, in the sense of vector optimizations.

The additive model we select is

$$\begin{aligned}
 & \max \sum_{i=1}^m g_i^- s_i^- + \sum_{r=1}^s g_r^+ s_r^+, \\
 & \text{subject to} \\
 & \sum_{j=1}^n x_{ij} \lambda_j + s_i^- = x_{io}, \quad i = 1, 2, \dots, m, \\
 & \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{ro}, \quad r = 1, 2, \dots, s, \\
 & \sum_{j=1}^n \lambda_j = 1, \\
 & \lambda_j, s_i^-, s_r^+ \geq 0.
 \end{aligned} \tag{2.15}$$

This model utilizes the “goal vector” approach of Thrall (1996a) in which the slacks in the objective are accorded “goal weights” which may be subjective or objective in character. Here we want to use these “goal weights” to ensure that the units of measure associated with the slack variables do not affect the optimal solution choices.

Employing the language of “dimensional analysis,” as in Thrall (1996a), we want these weights to be “contragredient” to insure that the resulting objective will be “dimensionless.” That is, we want the solutions to be free of the dimensions in which the inputs and outputs are stated. An example is the use of the input and output ranges in Cooper et al. (1999) to obtain  $g_i = 1/R_i^-$ ,  $g_r = 1/R_r^+$  where  $R_i^-$  is the range for the  $i$ th input and  $R_r^+$  is the range for the  $r$ th output. This gives each term in the objective of (2.15) a contragredient weight. The resulting value of the objective is dimensionless, as follows from the fact that the  $s_i^-$  and  $s_r^+$  in the numerators are measured in the same units as the  $R_i^-$  and  $R_r^+$  in the denominators. Hence the units of measure cancel.

The condition for efficiency given in Definition 1.3 in Chap. 1 for the CCR model is now replaced by the following simpler condition,

**Definition 2.1.** A DMU<sub>*o*</sub> evaluated by (2.15) is efficient if and only if all slacks are zero.

Thus, in the case of additive models it suffices to consider only condition (ii) in Definition 1.3. Moreover this condition emerges from the second stage solution procedure associated with the non-Archimedean  $\varepsilon > 0$  in (1.1). Hence we might expect that RTS characterizations will be related, as we see now.

To start our RTS analyses for these additive models we first replace the CCR projections of (2.3) with

$$\begin{aligned}
 \hat{x}_{io} &= x_{io} - s_i^{-*}, \quad i = 1, \dots, m, \\
 \hat{y}_{ro} &= y_{ro} + s_r^{+*}, \quad r = 1, \dots, s,
 \end{aligned} \tag{2.16}$$

where  $s_i^{-*}$  and  $s_r^{+*}$  are optimal slacks obtained from (2.15). Then we turn to the dual (multiplier) model associated with (2.15) which we write as follows,

$$\begin{aligned}
& \min \sum_{i=1}^m v_i x_{io} - \sum_{r=1}^s \mu_r y_{ro} + u_o, \\
& \text{subject to} \\
& \sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s \mu_r y_{rj} + u_o \geq 0, \quad j = 1, \dots, m, \\
& v_i \geq g_i^-, \mu_r \geq g_r^+; u_o \text{ free.}
\end{aligned} \tag{2.17}$$

We are thus in position to use Theorem 2.1 for “additive” as well as “radial measures” as reflected in the BCC and CCR models discussed in earlier parts of this chapter. Hence we again have recourse to this theorem where, however, we note the difference in objectives between (2.2) and (2.17), including the change from  $-u_o$  to  $+u_o$ . As a consequence of these differences we also modify (2.4) to the following,

$$\begin{aligned}
& \text{Maximize } \hat{u}_o, \\
& \text{subject to} \\
& \sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - \hat{u}_o \leq 0, \quad j = 1, \dots, m; j \neq 0, \\
& \sum_{r=1}^s \mu_r \hat{y}_{ro} - \sum_{i=1}^m v_i \hat{x}_{io} - \hat{u}_o = 0, \\
& \mu_r \geq g_r^+, v_i \geq g_i^-, \hat{u}_o \leq 0.
\end{aligned} \tag{2.18}$$

Here we have assumed that  $u_o^* < 0$  was achieved in a first-stage use of (2.17). Hence, if  $\hat{u}_o^* < 0$  is maximal in (2.18) then RTS are increasing at  $(\hat{x}_o, \hat{y}_o)$  in accordance with (i) in Theorem 2.1 whereas if  $\hat{u}_o^* = 0$  then (iii) applies and RTS are constant at this point  $(\hat{x}_o, \hat{y}_o)$  on the efficiency frontier.

For  $u_o^* > 0$  in stage one, the objective and the constraint on  $\hat{u}_o$  are simply reoriented in the manner we now illustrate by using (2.15) to evaluate  $E$  in Fig. 2.2 via

$$\begin{aligned}
& \max s^- + s^+, \\
& \text{subject to} \\
& \lambda_A + \frac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E + s^- = 4, \\
& \lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \frac{9}{2}\lambda_E - s^+ = \frac{9}{2}, \\
& \lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E = 1, \\
& s^-, s^+, \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E \geq 0,
\end{aligned}$$

where we have used unit weights for the  $g_i^-, g_r^+$ , to obtain the usual additive model formulation. (See Thrall (1996b) for a discussion of the applicable condition

for a choice of such “unity” weights.) This has an optimal solution with  $\lambda_C^* = \lambda_D^* = s^{-*} = 1/2$  and all other variables zero. To check that this is optimal we turn to the corresponding dual (multiplier) form for the above envelopment model, which is

$$\begin{aligned}
 & \min 4v - \frac{9}{2}\mu + u_o, \\
 & \text{subject to} \\
 & v - \mu + u_o \geq 0, \\
 & \frac{3}{2}v - 2\mu + u_o \geq 0, \\
 & 3v - 4\mu + u_o \geq 0, \\
 & 4v - 5\mu + u_o \geq 0, \\
 & 4v - \frac{9}{2}\mu + u_o \geq 0, \\
 & v, \mu \geq 1, \quad u_o \text{ free.}
 \end{aligned}$$

The solution  $v^* = \mu^* = u_o^* = 1$  satisfies all constraints and gives  $4v^* - (9/2)\mu^* + u_o^* = 1/2$ . This is the same value as in the preceding problem so that, by the dual theorem of linear programming, both solutions are optimal.

To determine the conditions for RTS, we use (2.16) to project  $E$  into  $E'$  with coordinates  $(\hat{x}, \hat{y}) = (7/2, 9/2)$  in Fig. 2.2. Then we utilize the following reorientation of (2.18),

$$\begin{aligned}
 & \min \hat{u}_o, \\
 & \text{subject to} \\
 & v - \mu + \hat{u}_o \geq 0, \\
 & \frac{3}{2}v - 2\mu + \hat{u}_o \geq 0, \\
 & 3v - 4\mu + \hat{u}_o \geq 0, \\
 & 4v - 5\mu + \hat{u}_o \geq 0, \\
 & \frac{7}{2}v - \frac{9}{2}\mu + \hat{u}_o = 0, \\
 & v, \mu \geq 1, \quad \hat{u}_o \geq 0.
 \end{aligned}$$

This also gives  $v^* = u^* = \hat{u}_o^* = 1$  so the applicable condition is (ii) in Theorem 2.1. Thus, RTS are decreasing at  $E'$ , the point on the BCC efficiency frontier which is shown in Fig. 2.2.

## 2.6 Multiplicative Models

The treatments to this point have been confined to “qualitative” characterizations in the form of identifying whether RTS are “increasing,” “decreasing,” or “constant.” There is a literature – albeit a relatively small one – which is directed to “quantitative” estimates of RTS in DEA. Examples are the treatment of scale elasticities in Banker et al. (1984), Førsund (1996) and Banker and Thrall (1992). However, there are problems in using the standard DEA models, as is done in these studies, to obtain scale elasticity estimates. Førsund (1996), for instance, lists a number of such problems. Also the elasticity values in Banker and Thrall (1992) are determined only within upper and lower bounds. This is an inherent limitation that arises from the piecewise linear character of the frontiers for these models. Finally, attempts to extend the Färe, Grosskopf and Lovell (1985, 1994) approaches to the determination of scale elasticities have not been successful. See the criticisms in Førsund (1996, p. 296) and Fukuyama (2000, p. 105). (Multiple output–multiple input production and cost functions, which meet the sub- and superadditivity requirements in economics, are dealt with in Panzar and Willig (1977). See also Baumol et al. (1982).)

This does not, however, exhaust the possibilities. There is yet another class of models referred to as “multiplicative models,” which were introduced by this name into the DEA literature in Charnes et al. (1982) – see also Banker et al. (1981) – and extended in Charnes et al. (1983) to accord these models nondimensional (=units invariance) properties such as those we have just discussed. Although not used very much in applications these multiplicative models can provide advantages for extending the range of potential uses for DEA. For instance, they are not confined to efficiency frontiers that are concave. They can be formulated to allow the efficiency frontiers to be concave in some regions and nonconcave elsewhere. See Banker and Maindiratta (1986). They can also be used to obtain “exact” estimates of elasticities in manners that we now describe.

The models we use for this discussion are due to Banker and Maindiratta (1986) – where analytical characterizations are supplied along with confirmation in controlled-experimentally designed simulation studies.

We depart from the preceding development and now use an output-oriented model, which has the advantage of placing this development in consonance with the one in Banker and Maindiratta (1986) – viz.,

$$\begin{aligned}
 & \max \gamma_o, \\
 & \text{subject to} \\
 & \prod_{j=1}^n x_{ij}^{\lambda_j} \leq x_{io}, \quad i = 1, \dots, m, \\
 & \prod_{j=1}^n y_{rj}^{\lambda_j} \geq \gamma_o y_{ro}, \quad r = 1, \dots, s, \\
 & \sum_{j=1}^n \lambda_j = 1, \\
 & \gamma_o, \lambda_j \geq 0.
 \end{aligned} \tag{2.19}$$

To convert these inequalities to equations we use

$$\begin{aligned}
 e^{s_i^-} \prod_{j=1}^n x_{ij}^{\lambda_j} &= x_{io}, \quad i = 1, \dots, m \\
 &\text{and} \\
 e^{-s_r^+} \prod_{j=1}^n y_{rj}^{\lambda_j} &= y_{ro}, \quad r = 1, \dots, s
 \end{aligned} \tag{2.20}$$

and replace the objective in (2.19) with  $\gamma e^{(\sum_{r=1}^s s_r^+ + \sum_{i=1}^m s_i^-)}$ , where  $s_i^-, s_r^+ \geq 0$  represent slacks. Employing (2.20) and taking logarithms we replace (2.19) with

$$\begin{aligned}
 \min -\tilde{\gamma}_o - \varepsilon \left( \sum_{r=1}^s s_r^+ + \sum_{i=1}^m s_i^- \right), \\
 \text{subject to} \\
 \tilde{x}_{io} &= \sum_{j=1}^n \tilde{x}_{ij} \lambda_j + s_i^-, \quad i = 1, \dots, m, \\
 \tilde{\gamma}_o + \tilde{y}_{ro} &= \sum_{j=1}^n \tilde{y}_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s, \\
 1 &= \sum_{j=1}^n \lambda_j, \\
 \lambda_j, s_r^+, s_i^- &\geq 0, \quad \forall j, r, i,
 \end{aligned} \tag{2.21}$$

where “ $\sim$ ” denotes “logarithm” so the  $\tilde{x}_{ij}$ ,  $\tilde{y}_{rj}$  and the  $\tilde{\gamma}_o$ ,  $\tilde{x}_{io}$ ,  $\tilde{y}_{ro}$  are in logarithmic units.

The dual to (2.21) is

$$\begin{aligned}
 \max \sum_{r=1}^s \beta_r \tilde{y}_{ro} - \sum_{i=1}^m \alpha_i \tilde{x}_{io} - \alpha_o, \\
 \text{subject to} \\
 \sum_{r=1}^s \beta_r \tilde{y}_{rj} - \sum_{i=1}^m \alpha_i \tilde{x}_{ij} - \alpha_o \leq 0, \quad j = 1, \dots, n, \\
 \sum_{r=1}^s \beta_r = 1, \\
 \alpha_i \geq \varepsilon, \beta_r \geq \varepsilon; \quad \alpha_o \text{ free in sign.}
 \end{aligned} \tag{2.22}$$

Using  $\alpha_i^*$ ,  $\beta_r^*$  and  $\alpha_o^*$  for optimal values,  $\sum_{r=1}^s \beta_r^* \tilde{y}_{ro} - \sum_{i=1}^m \alpha_i^* \tilde{x}_{io} - \alpha_o^* = 0$  represents a supporting hyperplane (in logarithmic coordinates) for DMU<sub>o</sub>, where efficiency

is achieved. We may rewrite this log-linear supporting hyperplane in terms of the original input/output values:

$$\prod_{r=1}^s y_{ro}^{\beta_r^*} = e^{\alpha_o^*} \prod_{i=1}^m x_{io}^{\alpha_i^*}. \quad (2.23)$$

Then, in the spirit of Banker and Thrall (1992), we introduce

**Theorem 2.7.** Multiplicative Model RTS,

- (i) RTS are increasing if and only if  $\sum \alpha_i^* > 1$  for all optimal solutions to (2.23).
- (ii) RTS are decreasing if and only if  $\sum \alpha_i^* < 1$  for all optimal solutions to (2.23).
- (iii) RTS are constant if and only if  $\sum \alpha_i^* = 1$  for some optimal solutions to (2.23).

To see what this means we revert to the discussion of (2.11) and introduce scalars  $a, b$  in  $(aX_o, bY_o)$ . In conformance with (2.23) this means

$$e^{\alpha_o^*} \prod_{i=1}^m (ax_{io})^{\alpha_i^*} = \prod_{r=1}^s (by_{ro})^{\beta_r^*}, \quad (2.24)$$

so that the thus altered inputs and outputs satisfy this extension of the usual Cobb–Douglas types of relations.

The problem now becomes: given an expansion  $a > 1$ , contraction  $a < 1$ , or neither, i.e.,  $a = 1$ , for application to all inputs, what is the value of  $b$  that positions the solution in the supporting hyperplane at this point? The answer is given by the following

**Theorem 2.8.** If  $(aX_o, bY_o)$  lies in the supporting hyperplane then  $b = a^{\sum_{i=1}^m \alpha_i^*}$ .

*Proof.* This proof is adopted from Banker et al. (2004). Starting with the expression on the left in (2.25) we can write

$$e^{\alpha_o^*} \prod_{i=1}^m (ax_{io})^{\alpha_i^*} = a^{\sum_{i=1}^m \alpha_i^*} e^{\alpha_o^*} \prod_{i=1}^m x_{io}^{\alpha_i^*} = \frac{a^{\sum_{i=1}^m \alpha_i^*}}{b} \prod_{r=1}^s (by_{ro})^{\beta_r^*}, \quad (2.25)$$

by using the fact that  $\sum_{r=1}^s \beta_r^* = 1$  in (2.22) and  $e^{\alpha_o^*} \prod_{i=1}^m x_{io}^{\alpha_i^*} = \prod_{r=1}^s y_{ro}^{\beta_r^*}$  in (2.23).

Thus, to satisfy the relation (2.24) we must have  $b = a^{\sum_{i=1}^m \alpha_i^*}$  as the theorem asserts. ■

Via this Theorem, we have the promised insight into reasons why more than proportionate output increases are associated with  $\sum_{i=1}^m \alpha_i^* > 1$ , less than proportionate increases are associated with  $\sum_{i=1}^m \alpha_i^* < 1$  and CRS is the applicable condition when  $\sum_{i=1}^m \alpha_i^* = 1$ .

There may be alternative optimal solutions for (2.22) so the values for the  $\alpha_i^*$  components need not be unique. For dealing with alternate optima, we return to (2.19) and note that a necessary condition for efficiency is  $\gamma_o^* = 1$ .



For full efficiency we must also have all slacks at zero in (2.20). An adaptation of (2.3) to the present problem, therefore, gives the following:

$$\begin{aligned} \prod_{j=1}^n x_{ij}^{\lambda_{ij}^*} &= e^{-s_i^*} x_{io} = x'_{io}, \quad i = 1, \dots, m, \\ \prod_{j=1}^n y_{rj}^{\lambda_{rj}^*} &= e^{s_r^{+*}} \gamma_o^* y_{ro} = y'_{ro}, \quad r = 1, \dots, s \end{aligned} \quad (2.26)$$

and  $x'_{io}$ ,  $y'_{ro}$  are the coordinates of the point on the efficiency frontier used to evaluate  $DMU_o$ .

Thus, we can extend the preceding models in a manner that is now familiar. Suppose that we have obtained an optimal solution for (2.22) with  $\sum_{i=1}^m \alpha_i^* < 1$ . We then utilize (2.26) to form the following problem

$$\begin{aligned} &\max \sum_{i=1}^m \alpha_i, \\ &\text{subject to} \\ &\sum_{r=1}^s \beta_r \tilde{y}_{rj} - \sum_{i=1}^m \alpha_i \tilde{x}_{ij} - \alpha_o \leq 0, \quad j = 1, \dots, n; j \neq 0, \\ &\sum_{r=1}^s \beta_r \tilde{y}'_{ro} - \sum_{i=1}^m \alpha_i \tilde{x}'_{io} - \alpha_o = 0, \\ &\sum_{r=1}^s \beta_r = 1, \\ &\sum_{i=1}^m \alpha_i \leq 1, \\ &\alpha_i \geq \varepsilon, \beta_r \geq \varepsilon; \quad \alpha_o \text{ free in sign.} \end{aligned} \quad (2.27)$$

If  $\sum_{i=1}^m \alpha_i^* = 1$  in (2.27), then RTS are constant by (iii) of Theorem 2.8. If the maximum is achieved with  $\sum_{i=1}^m \alpha_i^* < 1$ , however, condition (ii) of Theorem 2.7 is applicable and RTS are decreasing at the point  $x'_{io}, y'_{ro}; i = 1, \dots, m; r = 1, \dots, s$ .

If we initially have  $\sum_{i=1}^m \alpha_i^* > 1$  in (2.22), we replace  $\sum_{i=1}^m \alpha_i^* \leq 1$  with  $\sum_{i=1}^m \alpha_i^* \geq 1$  in (2.27) and also change the objective to minimize  $\sum_{i=1}^m \alpha_i^*$ . If the optimal value is greater than one, then (i) of Theorem 2.7 is applicable and the RTS are increasing. On the contrary, if we attain  $\sum_{i=1}^m \alpha_i^* = 1$  then condition (iii) applies and RTS are constant.

Theorem 2.8 also allows us to derive pertinent scale elasticities in a straightforward manner. Thus, using the standard logarithmic derivative formulas for elasticities, we obtain the following:

$$\frac{d \ln b}{d \ln a} = \frac{a}{b} \frac{db}{da} = \sum_{i=1}^m \alpha_i^*. \quad (2.28)$$

Consisting of a sum of component elasticities, one for each input, this overall measure of elasticity is applicable to the value of the multiplicative expression with which  $DMU_o$  is associated.

The derivation in (2.28) holds only for points where this derivative exists. However, we can bypass this possible source of difficulty by noting that Theorem 2.8 allows us to obtain this elasticity estimate via

$$\frac{\ln b}{\ln a} = \sum_{i=1}^m \alpha_i^*. \quad (2.29)$$

Further, as discussed in Cooper et al. (1996), it is possible to extend these concepts to the case in which all of the components of  $Y_o$  are allowed to increase by at least the factor  $b$ . However, we cannot similarly treat the constant,  $a$ , as providing an upper bound for the inputs since mix alterations are not permitted in the treatment of RTS in economics. See Varian (1984, p. 20) for requirements of RTS characterizations in economics.

In conclusion, we turn to properties of units invariance for these multiplicative models. Thus, we note that  $\sum_{i=1}^m \alpha_i^*$  is units invariant by virtue of the relation expressed in (2.28). The property of units invariance is also exhibited in (2.29) since  $a$  and  $b$  are both dimension free. Finally, we also have the following.

**Theorem 2.9.** The model given in (2.19) and (2.20) is dimension free. That is, changes in the units used to express the input quantities  $x_{ij}$  or the output quantities  $y_{rj}$  in (2.19) will not affect the solution set or alter the value of  $\max \gamma_o = \gamma_o^*$ .

*Proof.* Let

$$\begin{aligned} x'_{ij} &= c_i x_{ij}, & x'_{io} &= c_i x_{io}, & i &= 1, \dots, m, \\ y'_{rj} &= k_r y_{rj}, & y'_{ro} &= k_r y_{ro}, & r &= 1, \dots, s, \end{aligned} \quad (2.30)$$

where the  $c_i$  and  $k_r$  are any collection of positive constants. By substitution in the constraints for (2.20), we then have

$$\begin{aligned} e^{\bar{s}_i} \prod_{j=1}^n x'_{ij} \lambda_j &= x'_{io}, & i &= 1, \dots, m, \\ e^{s_r^+} \prod_{j=1}^n y'_{rj} \lambda_j &= \gamma_o y'_{ro}, & r &= 1, \dots, s, \\ \sum_{j=1}^n \lambda_j &= 1, \quad \lambda_j \geq 0, & j &= 1, \dots, n. \end{aligned} \quad (2.31)$$

Utilization of (2.30), therefore, gives

$$\begin{aligned}
 e^{s_i^-} c_i \sum_{j=1}^n \lambda_j \prod_{j=1}^n x_{ij}^{\lambda_j} &= c_i x_{io}, & i = 1, \dots, m, \\
 e^{-s_r^+} k_r \sum_{j=1}^n \lambda_j \prod_{j=1}^n y_{rj}^{\lambda_j} &= \gamma_o k_r y_{ro}, & r = 1, \dots, s, \\
 \sum_{j=1}^n \lambda_j &= 1, \quad \lambda_j \geq 0, & j = 1, \dots, n.
 \end{aligned} \tag{2.32}$$

However,  $\sum_{j=1}^n \lambda_j = 1$ , so  $c_i \sum_{j=1}^n \lambda_j = c_i$  and  $k_r \sum_{j=1}^n \lambda_j = k_r \quad \forall i, r$ . Therefore, these constants, which appear on the right and left of (2.32), all cancel. Thus, all solutions to (2.31) are also solutions to (2.20) and vice versa. It follows that the optimal value of one program is also optimal for the other. ■

We now conclude our discussion of these multiplicative models with the following:

**Corollary to Theorem 2.9.** The restatement of (2.20) in logarithmic form yields a model that is translation invariant.

*Proof.* Restating (2.31) in logarithmic form gives

$$\begin{aligned}
 s_i^- + \sum_{j=1}^n (\tilde{x}_{ij} + \tilde{c}_i) \lambda_j &= \tilde{x}_{io} + \tilde{c}_i, & i = 1, \dots, m, \\
 -s_r^+ + \sum_{j=1}^n (\tilde{y}_{rj} + \tilde{k}_r) \lambda_j &= \tilde{y}_{ro} + \tilde{k}_r + \tilde{\gamma}_o, & r = 1, \dots, s, \\
 \sum_{j=1}^n \lambda_j &= 1, \quad \lambda_j \geq 0, & j = 1, \dots, n.
 \end{aligned} \tag{2.33}$$

Once more utilizing  $\sum \lambda_j = 1$  we eliminate the  $\tilde{c}_i$  and  $\tilde{k}_r$  on both sides of these expressions and obtain the same constraints as in (2.21). Thus, as before, the solution sets are the same and an optimum solution for one program is also optimal for the other – including the slacks. ■

## 2.7 Summary and Conclusion

Although we have now covered all of the presently available models, we have not covered all of the orientations in each case. Except for the multiplicative models, we have not covered output-oriented objectives for a variety of reasons. There are no real problems with the mathematical development, but further attention must be

devoted to how changes in input scale and input mix should be treated when all outputs are to be scaled up in the same proportions. See the discussion in Cooper et al. (1996).

As also noted in Cooper et al. (1996), the case of IRS can be clarified by using Banker's most productive scale size to write  $(X_o\alpha, Y_o\beta)$ . The case  $1 < \beta/\alpha$  means that all outputs are increased by at least the factor  $\beta$  and RTS are increasing as long as this condition holds. The case  $1 > \beta/\alpha$  has the opposite meaning – viz., no output is increasing at a rate that exceeds the rate at which all inputs are increased. Only for CRS do we have  $1 = \beta/\alpha$ , in which case all outputs and all inputs are required to be increasing (or decreasing) at the same rate so no mix change is involved for the inputs.

The results in this chapter (as in the literature to date) are restricted to this class of cases. This leaves unattended a wide class of cases. One example involves the case where management interest is centered on only subsets of the outputs and inputs. A direct way to deal with this situation is to partition the inputs and outputs of interest and designate the conditions to be considered by  $(X_o^I\alpha, X_o^N, Y_o^I\beta, Y_o^N)$  where  $I$  designates the inputs and outputs that are of interest to management and  $N$  designates those which are not of interest (for such scale returns studies). Proceeding as described in the present chapter and treating  $X_o^N$  and  $Y_o^N$  as “exogenously fixed,” in the spirit of Banker and Morey (1986), would make it possible to determine the situation for RTS with respect to the thus designated subsets. Other cases involve treatments with unit costs and prices as in FGL (1994) and Sueyoshi (1999).

The developments covered in this chapter have been confined to technical aspects of production. Our discussions follow a long-standing tradition in economics that distinguishes scale from mix changes by not allowing the latter to vary when scale changes are being considered. This permits the latter (i.e., scale changes) to be represented by a single scalar – hence the name. However, this can be far from actual practice, where scale and mix are likely to be varied simultaneously when determining the size and scope of an operation. See the comments by a steel industry consultant that are quoted in Cooper, Seiford, and Tone (2000, p. 130) on the need for reformulating this separation between mix and scale changes to achieve results that more closely conform to needs and opportunities for use in actual practice.

There are, of course, many other aspects to be considered in treating RTS besides those attended to in the present chapter. Management efforts to maximize profits, even under conditions of certainty, require simultaneous determination of scale, scope, and mix magnitudes with prices and costs known, as well as the achievement of the technical efficiency, which is always to be achieved with *any* set of positive prices and costs. The topics treated in this chapter do not deal with such price–cost information. Moreover, the focus is on ex post facto analysis of already effected decisions. This can have many uses, especially in the control aspects of management where evaluations of performance are required. Left unattended in this chapter, and in much of the DEA literature, is the ex ante (planning) problem of how to use this knowledge to determine how to blend scale and scope with mix and other efficiency considerations when effecting future-oriented decisions.

## Appendix

In this Appendix, we first present the FGL approach. We then present a simple RTS approach without the need for checking the multiple optimal solutions as in Zhu and Shen (1995) and Seiford and Zhu (1999) where only the BCC and CCR models are involved. This approach will substantially reduce the computational burden because it relies on the standard CCR and BCC computational codes (see Zhu (2009) for a detailed discussion).

To start, we add to the BCC and CCR models by the following DEA model whose frontier exhibits nonincreasing returns to scale (NIRS), as in Färe, Grosskopf and Lovell (FGL 1985, 1994)

$$\begin{aligned}
 \theta_{\text{NIRS}}^* &= \min \theta_{\text{NIRS}}, \\
 \text{subject to} \\
 \theta_{\text{NIRS}} x_{io} &= \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i = 1, 2, \dots, m, \\
 y_{ro} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, 2, \dots, s, \\
 1 &\geq \sum_{j=1}^n \lambda_j, \\
 0 &\leq \lambda_j, s_i^-, s_r^+ \quad \forall i, r, j.
 \end{aligned} \tag{2.34}$$

The development used by FGL (1985, 1994) rests on the following relation

$$\theta_{\text{CCR}}^* \leq \theta_{\text{NIRS}}^* \leq \theta_{\text{BCC}}^*,$$

where “\*” refers to an optimal value and  $\theta_{\text{NIRS}}^*$  is defined in (2.34), while  $\theta_{\text{BCC}}^*$  and  $\theta_{\text{CCR}}^*$  refer to the BCC and CCR models as developed in Theorems 2.3 and 2.4.

FGL utilize this relation to form ratios that provide measures of RTS. However, we turn to the following tabulation that relates their RTS characterization to Theorems 2.3 and 2.4 (and accompanying discussion). See also Färe and Grosskopf (1994), Banker et al. (1996b), and Seiford and Zhu (1999)

	FGL Model	RTS	CCR Model
Case 1	If $\theta_{\text{CCR}}^* = \theta_{\text{BCC}}^*$	Constant	$\sum \lambda_j^* = 1$
Case 2	If $\theta_{\text{CCR}}^* < \theta_{\text{BCC}}^*$ then		
Case 2a	If $\theta_{\text{CCR}}^* = \theta_{\text{NIRS}}^*$	Increasing	$\sum \lambda_j^* < 1$
Case 2b	If $\theta_{\text{CCR}}^* < \theta_{\text{NIRS}}^*$	Decreasing	$\sum \lambda_j^* > 1$

It should be noted that the problem of nonuniqueness of results in the presence of alternative optima is not encountered in the FGL approach (unless output-oriented as well as input-oriented models are used), whereas they do need to be coincided, as in Theorem 2.3. However, Zhu and Shen (1995) and Seiford and Zhu (1999) develop an alternative approach that is not troubled by the possibility of such alternative optima.

We here present their results with respect to Theorems 2.3 and 2.4 (and accompanying discussion). See also Zhu (2009).

	Seiford and Zhu (1999)	RTS	CCR Model
Case 1	If $\theta_{CCR}^* = \theta_{BCC}^*$	Constant	$\sum \lambda_j^* = 1$
Case 2	$\theta_{CCR}^* \neq \theta_{BCC}^*$		
Case 2a	If $\sum \lambda_j^* < 1$ in any CCR outcome	Increasing	$\sum \lambda_j^* < 1$
Case 2b	If $\sum \lambda_j^* > 1$ in any CCR outcome	Decreasing	$\sum \lambda_j^* > 1$

The significance of Seiford and Zhu's (1999) approach lies in the fact that the possible alternate optimal  $\lambda_j^*$  obtained from the CCR model only affect the estimation of RTS for those DMUs that truly exhibit CRS and have nothing to do with the RTS estimation on those DMUs that truly exhibit IRS or DRS. That is, if a DMU exhibits IRS (or DRS), then  $\sum_j^n \lambda_j^*$  must be less (or greater) than one, no matter whether there exist alternate optima of  $\lambda_j$ , because these DMUs do not lie in the MPSS region. This finding is also true for the  $u_o^*$  obtained from the BCC multiplier models.

Thus, in empirical applications, we can explore RTS in two steps. First, select all the DMUs that have the same CCR and BCC efficiency scores regardless of the value of  $\sum_j^n \lambda_j^*$  obtained from model (2.5). These DMUs are CRS. Next, use the value of  $\sum_j^n \lambda_j^*$  (in any CCR model outcome) to determine the RTS for the remaining DMUs. We observe that in this process we can safely ignore possible multiple optimal solutions of  $\lambda_j$ .

## References

- Banker RD. Estimating most productive scale size using Data Envelopment Analysis. *Eur J Oper Res.* 1984;17:35–44.
- Banker RD, Bardhan I, Cooper WW. A note on returns to scale in DEA. *Eur J Oper Res.* 1996a;88:583–5.
- Banker RD, Chang H, Cooper WW. Equivalence and implementation of alternative methods for determining returns to scale in Data Envelopment Analysis. *Eur J Oper Res.* 1996b;89:473–81.
- Banker R, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci.* 1984;30:1078–92.
- Banker RD, Charnes A, Cooper WW, Schinnar A. A bi-extremal principle for Frontier Estimation and Efficiency Evaluation. *Manag Sci.* 1981;27:1370–82.
- Banker RD, Maindiratta A. Piecewise loglinear estimation of efficient production surfaces. *Manag Sci.* 1986;32:126–35.

- Banker RD, Morey R. Efficiency analysis for exogenously fixed inputs and outputs. *Oper Res.* 1986;34:513–21.
- Banker RD, Thrall RM. Estimation of returns to scale using Data Envelopment Analysis. *Eur J Oper Res.* 1992;62:74–84.
- Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M. and Zhu, J, Returns to scale in different DEA models, *European Journal of Operational Research*, 2004;154:345–362.
- Baumol WJ, Panzar JC, Willig RD. Contestable markets. New York: Harcourt Brace Jovanovich; 1982.
- Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *Eur J Oper Res.* 1978;2:429–44.
- Charnes A, Cooper WW, Seiford LM, Stutz J. A multiplicative model for efficiency analysis. *Socioecon Plann Sci.* 1982;16:213–24.
- Charnes A, Cooper WW, Seiford LM, Stutz J. Invariant multiplicative efficiency and piecewise Cobb-Douglas envelopments. *Oper Res Lett.* 1983;2:101–3.
- Cooper WW, Park KS, Pastor JT. RAM: A range adjusted measure of efficiency. *J Product Anal.* 1999;11:5–42.
- Cooper WW, Seiford LM, Tone K. Data envelopment analysis: a comprehensive text with models, references and DEA-Solver Software applications. Boston: Kluwer Academic Publishers; 2000.
- Cooper WW, Thompson RG, Thrall RM. Extensions and new developments in DEA. *Ann Oper Res.* 1996;66:3–45.
- Färe R, Grosskopf S. Estimation of returns to scale using data envelopment analysis: a comment. *Eur J Oper Res.* 1994;79:379–82.
- Färe R, Grosskopf S, Lovell CAK. The measurement of efficiency of production. Boston: Kluwer Nijhoff Publishing; 1985.
- Färe R, Grosskopf S, Lovell CAK. Production frontiers. Cambridge: Cambridge University Press; 1994.
- Førsund FR. On the calculation of scale elasticities in DEA models. *J Product Anal.* 1996;7:283–302.
- Frisch RA. Theory of production. Dordrecht: D. Rieoel; 1964.
- Fukuyama H. Returns to scale and scale elasticity in Data Envelopment Analysis. *Eur J Oper Res.* 2000;125:93–112.
- Golany B, Yu G. Estimating returns to scale in DEA. *Eur J Oper Res.* 1994;103:28–37.
- Panzar JC, Willig RD. Economies of scale in multi-output production. *Q J Econ.* 1977; XLI:481–493.
- Seiford LM, Zhu J. An investigation of returns to scale under Data Envelopment Analysis. *Omega.* 1999;27:1–11.
- Sueyoshi T. DEA duality on returns to scale (RTS) in production and cost analyses: an occurrence of multiple solutions and differences between production-based and cost-based RTS estimates. *Manag Sci.* 1999;45:1593–608.
- Thrall RM. Duality, classification and slacks in DEA. *Ann Oper Res.* 1996a;66:109–38.
- Thrall RM. The lack of invariance of optimal dual solutions under translation invariance. *Ann Oper Res.* 1996b;66:103–8.
- Varian H. Microeconomic analysis. New York: W.W. Norton; 1984.
- Zhu J. Setting scale efficient targets in DEA via returns to scale estimation methods. *J Oper Res Soc.* 2000;51(3):376–8.
- Zhu J. Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets. 2nd ed. Boston: Springer Science; 2009.
- Zhu J, Shen Z. A discussion of testing DMUs' returns to scale. *Eur J Oper Res.* 1995;81:590–6.

<http://www.springer.com/978-1-4419-6150-1>

Handbook on Data Envelopment Analysis

(Eds.) W.W. Cooper; L.M. Seiford; J. Zhu

2011, XXV, 497 p. 51 illus., 21 in color. With online files/update., Hardcover

ISBN: 978-1-4419-6150-1