Scaling production and improving efficiency in DEA: an interactive approach

Wilhelm Rödder¹ · Andreas Kleine² · Andreas Dellnitz²

Abstract DEA models help a DMU to detect its (in-)efficiency and to improve activities, if necessary. Efficiency is only one economic aim for a decision-maker; however, up- or downsizing might be a second one. Improving efficiency is the main topic in DEA; the long-term strategy towards the right production size should attract our attention as well. Not always the management of a DMU primarily focuses on technical efficiency but rather is interested in gaining scale effects. In this paper, a formula for returns to scale (RTS) is developed, and this formula is even applicable for interior points of technology. Particularly, technical and scale inefficient DMUs need sophisticated instruments to improve their situation. Considering RTS as well as efficiency, in this paper, we give an advice for each DMU to find an economically reliable path from its actual situation to better activities and finally to most productive scale size (mpss), perhaps. For realizing this path, we propose an interactive algorithm, thus harmonizing the scientific findings and the interests of the management. Small numerical examples illustrate such paths for selected DMUs; an empirical application in theatre management completes the contribution.

Keywords Data envelopment analysis · Returns to scale · Efficiency · Upsizing/downsizing · mpss

Introduction

The relations between efficient vectors of inputs and outputs for a given technology picture production functions, Shephard (1970), the characteristics of which predominantly are determined by their substitutionalities, and their returns to scale (RTS). The homogeneity of such a production function depends on the change, i.e., reaction of outputs with radial changes of inputs. Production theory distinguishes three forms of such reactions:

- constant RTS,
- decreasing RTS, and
- increasing RTS.

There is a great variety of methods to estimate such functional dependencies, cf. Coelli et al. (2005).

Data envelopment analysis (DEA) as a nonparametric approach permits the approximation of the efficient boundary of technologies. This approximation takes place via gathered data of inputs and outputs for the so-called decision-making units (DMUs), that is, classical DEA; numerous theoretical papers and applications prove its value, see, for instance, Bashiri et al. (2013), Shokrollahpour et al. (2016) and Ziari (2016) or for a overwhelming survey Emrouznejad and Yang (2017).

Already in their pioneering work, the authors in Banker et al. (1984) also tackled the problem of the DMUs’ RTS. They proved the sign of a variable in the so-called multiplier form of input-oriented DEA to indicate the RTS situation—rather than the RTS measure—of a DMU, and they restricted their analyses to efficient rather than...
inconstant/non-decreasing/non-increasing RTS if the optimal vector $u = (u_1, \ldots, u_n)$ satisfies $u_i = 0$. These results were generalized in Banker and Thrall (1992) or Sahoo et al. (2016) for efficient production points with variable $u_i$'s, i.e., non-unique RTS situations.

The authors in Førsund (1996), Førsund and Hjalmarsson (2004), Førsund et al. (2007), and Fukuyama (2000) advanced the RTS context by models of neoclassical production theory and developed equations of scale elasticity for efficient and non-efficient DMUs; an overview and a discussion regarding the concept of RTS are provided in Tone and Sahoo (2003) or Jahanshahloo and Soleimani-Damaneh (2004). In the present paper, we give an elegant and comprehensible proof and an easy interpretation of such scale elasticity, for efficient and non-efficient DMUs. Generally speaking, there are several ways for classifying a DMU’s RTS situation in DEA—either envelopment or multiplier form driven ways. However, the RTS measure is based on an optimal solution of the multiplier form; in this sense, the above-mentioned authors provide equivalent representations of the respective formula for calculating such scale elasticities.

A DMU which is informed about its relative inefficiency wants to react by input reduction, output increase, or both: Classical DEA theory recommends to proceed against the boundary of technology; for some activity planning procedures cf. Du et al. (2010), Du and Liang (2012), Homayounfar et al. (2014), Zhang et al. (2015) and Tohid and Khodadadi (2013). DEA software supports such instructions and helps the user to take respective actions. In all these calculations, scale elasticities are not taken into account. However, the additional knowledge of scale elasticities should encourage the DMUs also to make use of scaling effects. Consequently, if a DMU wants to maintain its BCC efficiency, then for increasing RTS it should upsize its production as any increasing inputs result in disproportionally higher increase of outputs. Decreasing RTS rather recommends downsizing due to lower disproportionality. Therefore, each DMU pursues two objectives: improvement of efficiency and upsizing or downsizing production. Mostly if not always the way to technology boundary is hard to realize. Labour law or social restrictions might forbid such rigid alterations. Is there a more convenient way towards the right production size and efficiency? It is, and the hitherto necessary information is available even for interior points of technology.

For the combination of efficiency and scaling improvement, we develop an (input-oriented) interactive algorithm, thus harmonizing scientific findings and managerial interests of the DMU under consideration. Stepwise it improves its efficiency and its scale size and hence runs through a path from its original activity towards mpss, if possible. The DMU’s awareness of inefficiency sometimes demands unrealistic input reduction. Rather it—the DMU—might communicate its disposition for a more reasonable reduction and the algorithm should respect this information.

Not always realizing the path from the actual activity of the DMU to most productive scale size (mpss) is an easy job. An impressive example for a vulnerable theatre scenery illustrates this issue.

The paper is organized as follows. In the second section, we present preliminaries of DEA. In the next section, two forms of activity changes are given, one maintaining BCC efficiency and one maintaining CCR efficiency. The following section is dedicated to situations of non-unique RTS. The next section provides the central topic of this paper: an interactive and iterative algorithm towards mpss. In the following section, an empirical example illustrates the new method. For 30 German theatres, efficiencies and RTS are calculated, and for a selected theatre, the proposed method is outlined. The last section contains a short resume and delineates prospects of further research.

### Preliminaries

Koopmans’s activity analyses are the roots of DEA, cf. Koopmans (1951). Activities are processes which transform objects in other objects. If such objects are material or immaterial goods, such a process by definition is a production process, cf. Frisch (1965), p. 3. The set of all such processes is the production possibility set or technology $T$, for short. The activity over a certain time period transforms the input $x \in \mathbb{R}^M_+$ into the output $y \in \mathbb{R}^J_+$ and thus characterizes the performance of a DMU, such as a project, a corporation, or even a non-profit utility. Once a technology is determined, DEA theory allows for the efficiency measurement of any activity. We restrict our attention to input-oriented efficiency measures and omit output orientation, cf. Banker et al. (1984), however. Partial inefficiencies which might occur with radial input reduction will not be considered here, either; the reader is referred to Charnes and Cooper (1984) or again Banker et al. (1984).

The authors in Charnes et al. (1978), following Debreu (1951), and Farell (1957) developed linear optimization problems (LOPs) measuring efficiency.

For each DMU, $k$ solve

$$
\begin{align}
\min \theta_k \\
\text{s.t.:} \quad & \theta_k x_k - \sum_j \lambda_{kj} y_j \geq 0 \\
& \sum_j \lambda_{kj} x_j \geq y_k \\
& \lambda_{kj} \geq 0 \quad \forall j, \quad \theta_k \text{ free}
\end{align}
$$

and --- for constant returns to scale (CRS) (1.1)
\[
\sum_\lambda \nu_{\lambda j} = 1 \quad \text{for variable returns to scale (VRS).} \quad (1.2)
\]

(1.1) and (1.2) are called \textit{envelopment form} of DEA. If the optimal \( \theta_k \) is equal to 1 DMU, \( k \) is efficient and inefficient, otherwise.

The dual problems of (1.1) and (1.2) are (2.1) and (2.2), again LOPs.

\[
\begin{align*}
\max g_k &= u_k^T y_k + u_k \\
\text{s.t.:} & \quad v_k^T x_k = 1 \\
& \quad u_k^T y_j + u_k - v_k^T x_j \leq 0 \quad \forall j \\
& \quad u_k, v_k \geq 0 \\
\end{align*}
\quad (2.1)
\]

and

\[
\begin{align*}
\max u_k &= 0 \quad \text{for CRS} \\
\text{s.t.:} & \quad v_k^T x_k = 1 \\
& \quad u_k^T y_j + u_k - v_k^T x_j \leq 0 \quad \forall j \\
& \quad u_k, v_k \geq 0 \\
\end{align*}
\quad (2.2)
\]

Such weights \( v_k \) and \( u_k \) of inputs and outputs often are called virtual prices. Neither are they preassigned nor market-based, they just suit a suitable means for each DMU \( k \) to stress its own efficiency. This attitude is called ‘self-appraisal’. Already in their pioneering work, Banker et al. (1984) demonstrate that for an efficient activity \((x_k, y_k)\), the optimal value of \( u_k \) in (2.2) indicates constant, non-decreasing, or non-increasing RTS, though the authors refer to the sign of the respective variable \( u_k \), only, and do not fully exploit the information from an optimal solution \( u_k^*, v_k^*, u_k^*, g_k^* \). More on that in the next section.

In the DEA literature, models (1.1), (2.1) and (1.2), (2.2) often are named by the acronyms CCR and BCC, due to their creators Charnes, Cooper, Rhodes and Banker, Charnes, Cooper, respectively. Whenever convenient, we follow such practice.

\section*{Activity changes}

\section*{Activity change under constant BCC efficiency}

Let \( u_k^*, v_k^*, u_k^*, g_k^* \) be an optimal solution of (2.2). Then, we have

\[
\frac{u_k^T y_k + u_k}{v_k^T x_k} = g_k^*.
\]

and we assume the optimal solution to be unique, first, cf. Banker et al. (1984), p. 1086. Multiplying by the denominator and reordering terms yield the following equation:

\[
\begin{align*}
&u_k^T y_k + u_k - g_k^* v_k^T x_k = 0. \\
\end{align*}
\quad (3)
\]

\textbf{Theorem 1} \quad For the equation \( u_k^T y_k + u_k - g_k^* v_k^T x_k = 0 \)
and a radial change \( x_k \rightarrow (1+\delta)x_k \) and a radial change \( y_k \rightarrow (1+\epsilon_k)y_k \), the equation \( u_k^T (1+\epsilon_k)y_k + u_k - g_k^* v_k^T (1+\delta)x_k = 0 \) holds iff
\[
\epsilon_k = \delta \frac{u_k^T y_k + u_k}{u_k^T y_k}. \\
\quad (4)
\]

\textbf{Proof}

\[
\begin{align*}
&u_k^T (1+\epsilon_k)y_k + u_k - g_k^* v_k^T (1+\delta)x_k = 0 \\
\iff & (1+\delta) \left[ u_k^T y_k + u_k - g_k^* v_k^T x_k \right] + u_k^T (1+\epsilon_k)y_k = 0 \\
\iff & \delta u_k^T y_k + u_k^T \epsilon_k y_k = u_k^T y_k + \delta \left( u_k^T y_k + u_k^T \epsilon_k y_k \right) \\
\iff & \delta \frac{u_k^T y_k + u_k}{u_k^T y_k} = \epsilon_k.
\end{align*}
\]

\[\square\]

Theorem 1 determines the necessary radial change of output with a radial change of input so as to maintain efficiency \( g_k^* \) constant. Equation (4) provides the respective relation between input/output changes. Please note that

\[
\epsilon_k = \delta \frac{u_k^T y_k + u_k}{u_k^T y_k} = \delta \frac{g_k^*}{g_k^* - u_k^*}.
\]

Therefore, the comprehensible result of theorem 1 corresponds to the well-known scale elasticity measure of F{"o}rsund et al. (2007). From this equation and from Eq. (4), it is obvious which role \( u_k^* \) plays in this context.

\section*{Conclusion 1}

\begin{itemize}
  \item \( |\epsilon_k| = |\delta| \) for \( u_k^* = 0 \): \( y \) changes radially to the same amount as \( x \) does \iff \text{constant RTS}
  \item \( |\epsilon_k| > |\delta| \) for \( u_k^* > 0 \): \( y \) changes radially to a greater amount than \( x \) does \iff \text{increasing RTS}
  \item \( |\epsilon_k| < |\delta| \) for \( u_k^* < 0 \): \( y \) changes radially to a smaller amount than \( x \) does \iff \text{decreasing RTS}.
\end{itemize}

All statements are valid for an efficient \( (g_k^* = 1) \) as well as for an inefficient \( (g_k^* < 1) \) DMU \( k \).

Perhaps, the second bullet point needs some attention. Equation (3) is valid at (in-)efficiency level \( g_k^* \); execution of input projection \( \bar{x}_k = g_k^* x_k \) results in

\[
\begin{align*}
&u_k^T y_k + u_k^* - g_k^* v_k^T x_k = 0. \\
\end{align*}
\quad (3')
\]

RTS for this equation obviously is the same as for (3). If \( g_k^* < 1 \), (3) characterizes RTS (directly) for an interior point of technology, see also Dellnitz (2016), whereas (3') determines the same RTS for the respective boundary point; see also, e.g., Fukuyama (2000). In the remainder of this contribution, we make use of this fact: moving in the interior of technology is possible without projection on the boundary!
The measure of RTS for an activity \((\mathbf{x}_k, \mathbf{y}_k)\) is \(u_k^T \mathbf{y} + u_r^* - g^*_k \mathbf{y}_k^T \mathbf{x} = 0\), rather than \(u_r^*\). This measure is not only a function of \(u_r^*\), but also of prices \(u_k^*\) and outputs \(\mathbf{y}_k\). For equivalent equations, confer again Førsund et al. (2007). The authors in Podinovski et al. (2009) grabbed this question again and simplified respective hyperplane in and outputs rather than the fix activity of DMU \(k\). The parameter \(u_k^*\) on the right-hand side:

\[
\min \theta_k \\
\text{s.t.: } \theta_k \mathbf{x}_k - \sum_j \lambda'_{kj} \cdot \mathbf{x}_j \geq 0 \\
\sum_j \lambda'_{kj} \cdot \mathbf{y}_j \geq \mathbf{y}_k \\
\sum_j \lambda_{kj} = r \\
\lambda_{kj} \geq 0 \quad \forall j, \quad \theta_k \text{ free.}
\]

Interesting enough, the parametrization of activity \((\mathbf{x}_k, \mathbf{y}_k)\) by \((\mathbf{x}_k/r, \mathbf{y}_k/r)\) is equivalent to reciprocal sensitivity analysis of the right-hand side of the convexity restriction in (1.2); let \([r^-, r^+]\) be the stability region. Activity change beyond these limits would make the input projection of \((\mathbf{x}_k/r, \mathbf{y}_k/r)\) fall off the actual facet of technology. Running on a trajectory \((\mathbf{x}_k, \mathbf{y}_k)/r\) lets DMU \(k\)’s CCR efficiency unchanged, of course. Note that \(r > 1\) implies in downsizing and \(r < 1\) in up-sizing DMU \(k\)’s activity.

In this section, two forms of activity changes were proposed:

- maintaining BCC efficiency in Sect. “Activity change under constant BCC efficiency” and
- maintaining CCR efficiency in Sect. “Activity change under constant CCR efficiency”.

How to combine these transformations for creating a path towards mpss, i.e., improving BCC efficiency and CCR efficiency or productivity, respectively, is topic of the algorithmic interactive approach in Sect. “Improving scale size and efficiency: an interactive approach”.

Before doing so, however, we need some statements on non-unique RTS. Sometimes a DMU must decide under aggravated conditions whether it should expand or reduce activity. In those cases, non-unique RTS play a decisive role.

### Non-unique RTS

The authors in Banker and Thrall (1992), Golany and Yu (1997) study the consequences of a non-unique solution \(u_k^*, v_k^*, u_r^*, g^*_k\) of problem (2.2), and specially focus on the non-uniqueness of \(u_k^*\). The former authors seek for geometrical characterizations of efficiency points and then calculate intervals for \(u_k^*\), cf. Banker and Thrall (1992), p. 82 equations (7) and (8). We follow their reasoning and generalize. Solve again
Let \( g_k^* \) and solve (9), cf. Mardani Shahrbabak and Noura (2011):

\[
\begin{align*}
  u_k^+ & = \sup \ u_k \text{ and } u_k^- = \inf \ u_k \\
  \text{s.t.:} \quad & v_k^T x_k = 1 \\
  & u_k^T y_k + u_k^- - v_k^T x_k \leq 0 \quad \forall j \\
  & u_k, v_k \geq 0 \quad \text{and } u_k \text{ free.}
\end{align*}
\]

Apply \( g_k^* \) and solve (9), cf. Mardani Shahrbabak and Noura (2011):

\[
\begin{align*}
  u_k^+ & = \sup \ u_k \text{ and } u_k^- = \inf \ u_k \\
  \text{s.t.:} \quad & v_k^T x_k = 1 \\
  & u_k^T y_k + u_k^- - v_k^T x_k \leq 0 \quad \forall j \\
  & u_k, v_k \geq 0 \quad \text{and } u_k \text{ free.}
\end{align*}
\]

Let \( u_k^+, v_k^+, u_k^-, v_k^- \) and \( u_k^+, v_k^- \) be the corresponding optimal solutions of (9). Equation (4) of Corollary 1 allows for the following generalization:

\[
\begin{align*}
  e_k^+ & = \delta \frac{u_k^T y_k + u_k^+}{u_k^+ y_k} \quad \text{and} \quad e_k^- = \delta \frac{u_k^T y_k + u_k^-}{u_k^- y_k}.
\end{align*}
\]

The numerator in (10) in each case equals \( g_k^* \), but is composed differently. The respective changes from \((x_k, y_k)\) to \((1 + \delta)x_k\) and \((1 + \epsilon_k)\) \(y_k\) must be feasible in \( T \), of course.

**Conclusion 2** If \( 0 < u_k^- \leq u_k \leq u_k^+ \), then for all \( u_k \) increasing RTS prevail at \((x_k, y_k)\); if \( u_k^- \leq u_k \leq u_k^+ < 0 \), then decreasing RTS. The case \( u_k^- \leq 0 \leq u_k^+ \) yields decreasing, increasing and specially constant RTS. This property remains valid for all activities \((x, y) \in T\) located on the hyperplane at efficiency level \( g_k^* \) like in (5).

**Example** For the activities of 7 DMUs, we have the data in Table 1. Table 2 contains the optimal weights, efficiencies, and RTS of the DMUs; the shaded area in Fig. 1 shows the respective technology. In addition, Fig. 1 illustrates non-unique RTS.

It shows that DMU 1 under model (2.2) performs with efficiency \( g_1^* = 1 \). Problem (9) yields \( u_1^- = \frac{3}{8} \) and \( u_1^+ = 1 \). For the equation \( 1 = u_1^1 + u_1^- \cdot v_1 \cdot 2 = 0 \), respectively, we consider the solutions

\[
\begin{align*}
  & u_1 = \frac{7}{8}, \quad u_1 = \frac{1}{8}, \quad v_1 = \frac{1}{2}; \\
  & u_1 = 1, \quad u_1 = 0, \quad v_1 = \frac{1}{2}; \\
  & u_1 = 0, \quad u_1 = 1, \quad v_1 = \frac{1}{2}.
\end{align*}
\]

The second case is degenerated and the third one violates the inequalities in (9) for DMU 2 to DMU 5. The only hyperplane which permits activity changes, feasible in \( T \), is the one with \( u_1 = \frac{3}{8} \). It contains DMU 2, ditto with \( g_2^* = 1 \).

DMU 6 under model (2.2) performs with efficiency \( g_6^* = \frac{3}{4} \). Problem (9) yields a unique \( u_6^- = u_6^+ = \frac{7}{12} \). The dashed line captioned with \( u_6 = \frac{7}{12} \) in Fig. 1 shows the hyperplane of DMU 6 at efficiency level \( g_6^* \). From equation \( g_6^* = \frac{u_k^2 + u_k^*}{y_k^3} \) or \( u_6 \cdot 2 + u_6 = \frac{3}{4} \cdot v_6 \cdot 3 = 0 \), respectively, we get: \( u_6 = \frac{1}{12} \), \( v_6 = \frac{5}{12} \). The hyperplane of DMU 6 which is illustrated in Fig. 1 is the equation \( \frac{1}{12} y + \frac{7}{12} - \frac{3}{4} \cdot \frac{1}{2} x = 0 \). Note that the slope of this equation is 3—whereas with 0.1, \( \frac{7/12 + 7/12}{1/12} = 0.45 \), the RTS is 45% for a 10% increase. DMU 6 operates under strictly increasing RTS.

---

**Table 1** Inputs/outputs of seven DMUs

|        | DMU 1 | DMU 2 | DMU 3 | DMU 4 | DMU 5 | DMU 6 | DMU 7 |
|--------|-------|-------|-------|-------|-------|-------|-------|
| Input  | 2     | \( \frac{1}{2} \) | 5     | 8     | 8     | 3     | 3     |
| Output | 1     | 3     | 7     | 9     | 8     | 2     | \( \frac{1}{2} \) |

**Table 2** Efficiency, optimal weights, and RTS of seven DMUs

|  | \( g_k^* \) | \( v_k^+, u_k^+, u_k^- \) | \( v_k^+, u_k^-, u_k^- \) | \( e_k^* = \delta \frac{u_k^T y_k + u_k^+}{u_k^+ y_k} \) | \( e_k^- = \delta \frac{u_k^T y_k + u_k^-}{u_k^- y_k} \) |
|---|-------------|-----------------|-----------------|-----------------|-----------------|
| DMU 1 | 1 | \( \frac{1}{2}, 0, 1 \) | \( \frac{1}{2}, \frac{3}{8}, \frac{1}{8} \) | \( e_1^+ = \delta \cdot \infty \) | \( e_1^- = \delta \cdot 8 \) |
| DMU 2 | 1 | \( \frac{3}{8}, \frac{1}{10}, \frac{7}{12} \) | \( \frac{3}{8}, \frac{1}{2}, \frac{1}{4} \) | \( e_2^+ = \delta \cdot 0 \) | \( e_2^- = \delta \cdot \frac{1}{4} \) |
| DMU 3 | 1 | \( \frac{1}{2}, \frac{1}{10}, \frac{1}{8} \) | \( \frac{1}{2}, \frac{1}{2}, -1 \) | \( e_3^+ = \delta \cdot \frac{1}{2} \) | \( e_3^- = \delta \cdot \frac{1}{2} \) |
| DMU 4 | 1 | \( \frac{1}{8}, \frac{3}{8}, - \frac{3}{4} \) | \( \frac{1}{8}, \infty, - \infty \) | \( e_4^+ = \delta \cdot \frac{1}{8} \) | \( e_4^- = \delta \cdot \frac{1}{8} \) |
| DMU 5 | \( \frac{13}{16} \) | \( \frac{1}{8}, \frac{3}{16}, - \frac{11}{16} \) | \( \frac{1}{8}, \frac{3}{16}, - \frac{11}{16} \) | \( e_5^+ = \delta \cdot \frac{13}{16} \) | \( e_5^- = \delta \cdot \frac{1}{16} \) |
| DMU 6 | \( \frac{2}{3} \) | \( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \) | \( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \) | \( e_6^+ = \delta \cdot \frac{2}{3} \) | \( e_6^- = \delta \cdot \frac{2}{3} \) |
| DMU 7 | \( \frac{2}{15} \) | \( \frac{1}{3}, 0, \frac{2}{3} \) | \( \frac{1}{3}, 0, \frac{2}{3} \) | \( e_7^+ = \delta \cdot \infty \) | \( e_7^- = \delta \cdot \infty \) |
Improving scale size and efficiency: an interactive approach

From Eqs. (2.2) and (4), DMU $k$ calculates its relative efficiency and its RTS. Even if the respective activity is an interior point of technology, the DMU knows the exact rate of radial output/input change to keep BCC efficiency constant, cf. Eq. (3’) and the subsequent text. But what must be done to improve productivity and simultaneously make the right scaling decision? After all, Eq. (4) is a recommendation of upsizing or downsizing production, see Conclusion 1. However, how far should this production sizing go. Banker (1984) formulated the concept of mpss activity. An activity has most productive scale size, when CCR and BCC efficiency coincide and are equal to 1. The author showed a way for each DMU how to reach this goal in one step without taking interactive communication with a DMU into account.

Here, the situation is different: How can even an inefficient DMU find its way stepwise to mpss avoiding this lack? Before giving a general answer to this question, we study the example again.

Example (continued).

First, we focus on DMU 5. It might pursue two objectives

- downsize activities due to decreasing RTS; as $u_5^x = u_5^y = -\frac{11}{16} < 0$,

- improve efficiency from actual $g_5^y = \frac{13}{16}$.

The two arrows in Fig. 2 give an idea of possible activity changes.

If DMU 5 runs on the dotted line, it realizes downsizing on its (in-)efficiency hyperplane, i.e., with constant BCC efficiency. It should stop when its input projection falls on another facet of technology and check its situation with respect to new weights.

So far downsizing; and next the goal efficiency improvement. We propose an interactive step. DMU 5 must find out its input reduction potential: What is the minimal input portion just to meet the benchmark $\tilde{y}_5$? Choose a number $g_5^y \leq f_5 \leq 1$, such that $\tilde{x}_5 = \hat{x}_5 \cdot f_5$ suffices to produce $\tilde{y}_5$. Make $(\tilde{x}_5, \tilde{y}_5)$ the new activity. For $f_5 = g_5^y$, DMU 5 becomes efficient, and for $f_5 = 1$, it has no input reduction potential at all. For $g_5^y < f_5 < 1$, it remains inefficient, but improves efficiency from $g_5^y$ to $f_5$. In Fig. 2, $f_5 = g_5^y$ even makes it a mpss.

Now, consider DMU 6. A first iteration step concerning upsizing yields input $\tilde{x}_6 = \frac{10}{7}$ and output $\tilde{y}_6 = 3$. For an exemplary demonstration, we assume a radial input reduction factor $f_6 = \frac{17}{20}$ resulting in $\tilde{x}_6 = \hat{x}_6 \cdot \frac{17}{20} = \frac{17}{6}$ and $\tilde{y}_6 = \tilde{y}_6 = 3$. This activity is a start point for a second iteration. A second upsizing results in $\tilde{x}_6 = \frac{17}{6}$ and $\tilde{y}_6 = 7$, and a second reduction step with $f_6 = \frac{15}{17}$ yields $(\tilde{x}_6 = 5, \tilde{y}_6 = 7)$. In addition, this makes DMU 6 mpss.
Whether or not DMU 6 finally reaches mpss obviously depends on its readiness for respective input reductions. In this case, the iterative procedure results in scale efficiency 1 and even mpss.

Running on BCC efficiency hyperplanes has two flaws:

1. \(u^T \mathbf{y}_k\) might be 0 and this makes \(u^T \mathbf{y}_k + u^T \mathbf{y}_k\) indefinite, cf. Theorem 1. Consider DMU 7: Its input projection falls on the vertical facet with \(u^*_7 = 0\). Consequently, RTS becomes indefinite, and hence, Eq. (6) is undefined. In such cases, the ideal path on BCC-(in-)efficiency hyperplanes, as was demonstrated for DMUs 5 and 6, is blocked.

2. Assume that DMU 7—besides the above-mentioned problem—succeeded in realizing activity (3, 1), see Fig. 1. Its efficiency remains \(\frac{2}{3}\), but now, Eq. (6) permits different weight systems for this efficiency. Which of these weight systems DMU 7 should select for further activity change remains an open question. In addition, this problem worsens, the more BCC efficiency hyperplanes pass through (3, 1). For high-dimensional DEA, this causes a severe problem.

We overcome these flaws by a three-step procedure.

1. Solve (1.2) with optimal value of objective function \(\theta_k\).
2. Parameterize the 1 in the convexity restriction of (1.2). Let \([r^-, r^+]\) be the range of this sensitivity analysis. Make \((\hat{x}_k, \hat{y}_k)^1 = (x_k, y_k)/r^\pm\) like in (8) and respective comments. Mind the fact that the scaling direction, \(r^-\) or \(r^+\), depends on the RTS, including \(\infty\) and 0, cf. Table 2.
3. Solve (1.2) for \((\hat{x}_k, \hat{y}_k)\) with optimal efficiency \(\hat{\theta}_k\). Make \((\hat{x}_k, \hat{y}_k) = (x_k \cdot \hat{\theta}_k/\hat{\theta}_k, y_k)\). This step traces \((\hat{x}_k, \hat{y}_k)\) back to BCC-(in-)efficiency hyperplane of DMU \(k\) at level \(\hat{\theta}_k\).

The following algorithm formalizes our explanations so far; for the stop criterion, we need both, the optimal CCR and BCC efficiency: \(g_k^{CR}\) and \(g_k^{BCC}\).

**Algorithm [MPSS]:**

For a DMU \(k \in 1, \ldots, J\)
1. solve (2.1) with solution \(g_k^{CR}\) and (2.2) with \(g_k^{BCC}\)
2. solve (9) with optimal solutions \(v^+_k, u^+_k, u^-_k, v^-_k\)
3. **Do until** \(g_k^{CR*} = g_k^{BCC}\)
4. solve (1.2) with \(\theta_k\)
5. if \(0 < u^-_k \leq u^+_k\) (increasing RTS) then
6. parameterize 1 in (1.2) with result \(r^-\)
7. set \(r = r^-\)
8. elseif \(u^-_k \leq u^+_k < 0\) (decreasing RTS) then
9. parameterize 1 in (1.2) with result \(r^+\)
10. set \(r = r^+\)
11. **end if**
12. make \((\hat{x}_k, \hat{y}_k) = (x_k, y_k)/r\)
13. solve (1.2) for \((\hat{x}_k, \hat{y}_k)\) with optimal efficiency \(\hat{\theta}_k\)
14. make \((\hat{x}_k, \hat{y}_k) = (x_k \cdot \hat{\theta}_k/\hat{\theta}_k, y_k)\)
15. choose \(x'_k = x_k \cdot \hat{\theta}_k\) with \(\hat{\theta}_k \leq \hat{\theta}_k\) \(\leq 1\)
16. replace \((x_k, y_k)\) by \((\hat{x}_k, \hat{y}_k)\)
17. for \((x_k, y_k)\) solve again (2.1) with optimal solution \(g_k^{CR*}\)
18. set \(g_k^* = \phi_k^\pm\)
19. **Loop**

To make the steps of the algorithm more transparent, we illustrate the first two iteration steps of DMU 7 in Fig. 3. DMU 7 is of particular interest, because its input projection has an infinite scale elasticity. Therefore, we apply an activity change under constant CCR efficiency like in Sect. “Activity change under constant CCR efficiency” (step 12). Then, step 14 traces back this activity to the prior BCC efficiency. Step 15 reflects the DMU’s disposition for input reduction. The continuing process of activity changes as indicated in Fig. 3 follows the same reasoning.

This algorithm is only one possible stepwise approximation of mpss, of course. So the order of scale-sizing and efficiency improvement might change, so up- and downsizing might be decomposed in smaller layers, etc. However, each specification of such an algorithm must comprise

- up- or downsizing with constant efficiency,
- interactive efficiency improvement,

in order to pursue the DMU’s two objectives.

So far the algorithm. How a DMU in the medium or in the long term can realize such alterations of its activities is due to its economical environment and its change management. These questions are beyond of the scope of this paper.

---

Fig. 3 Two steps of the proposed algorithm
Interactive improvement of activities in theatres

A rich theatre scenery is considered the basis of a broad cultural supply for the society, worldwide. The German theatre landscape follows this tradition. Therefore, in the season 2013/14, 142 public theatres attracted 21 million visitors offering more than 67,000 events.

Nevertheless, in times of budgets in deficit in almost all countries, the cost performance ratio of such activities is a major challenge for local authorities and politicians. There are some studies about this question, in which DEA plays a decisive role, cf. Tobias (2003), Marco-Serrano (2006), Kleine and Hoffmann (2013). What are desired outputs of a theatre, what the respective inputs, and for which homogenous subgroup of all theatres does such an efficiency measurement make sense. For our purpose, we consider two inputs:

- number of seats
- personnel expenses (million euro)

and three outputs: number of

- events
- productions
- visitors

as the database for DEA among 30 so-called three division theatres which offer drama, music, and dance performances. In Table 3, we show inputs and outputs for the season 2013/14—data from the German Stage Association, Deutscher Bühnenverein (2013)—together with respective CCR and BCC efficiencies plus RTS characteristics—IRS, CRS, and DRS—for these 30 theatres. All findings are results of standard DEA software. Note that each theatre is labelled by its city code.

Table 3 Inputs/outputs, efficiencies, and RTS of 30 theatres

| DMU | Seats | Pers. exp. | Events | Prod. | Visitors | $\hat{g}_k^*$ | $\hat{g}_k^*$ | RTS  |
|-----|-------|------------|--------|-------|----------|-------------|-------------|------|
| BI  | 3622  | 18.351     | 631    | 35    | 221,051  | 0.8940      | 1           | DRS  |
| C   | 2792  | 24.683     | 785    | 75    | 193,889  | 0.8370      | 1           | DRS  |
| CB  | 1236  | 17.791     | 431    | 43    | 121,126  | 0.7760      | 0.7810      | DRS  |
| CO  | 1661  | 11.495     | 413    | 37    | 88,178   | 0.8350      | 0.9040      | IRS  |
| DA  | 1674  | 27.752     | 644    | 51    | 236,710  | 0.8640      | 0.8860      | DRS  |
| DE  | 1645  | 15.476     | 922    | 51    | 150,798  | 1           | 1           | CRS  |
| DO  | 2856  | 31.868     | 672    | 54    | 231,292  | 0.6200      | 0.6400      | DRS  |
| G   | 2049  | 15.277     | 759    | 59    | 135,465  | 1           | 1           | CRS  |
| GI  | 1098  | 11.693     | 432    | 40    | 117,475  | 1           | 1           | CRS  |
| H   | 2361  | 49.844     | 1261   | 81    | 409,431  | 1           | 1           | CRS  |
| HBV | 803   | 12.006     | 532    | 37    | 128,535  | 1           | 1           | CRS  |
| HD  | 4831  | 19.143     | 1144   | 48    | 212,407  | 1           | 1           | CRS  |
| HZ  | 940   | 6.843      | 243    | 21    | 94,447   | 1           | 1           | CRS  |
| KA  | 1923  | 39.899     | 971    | 70    | 315,162  | 0.9660      | 0.9910      | DRS  |
| KL  | 3744  | 25.201     | 807    | 46    | 225,099  | 0.7160      | 0.7700      | DRS  |
| HZ  | 933   | 16.106     | 485    | 30    | 119,520  | 0.7960      | 0.8610      | IRS  |
| KO  | 3208  | 12.432     | 270    | 25    | 85,860   | 0.5790      | 0.6220      | IRS  |
| KS  | 3812  | 29.598     | 234    | 47    | 222,660  | 0.6350      | 0.6960      | DRS  |
| MD  | 1455  | 22.422     | 895    | 67    | 172,226  | 0.9990      | 1           | DRS  |
| MS  | 1230  | 13.104     | 569    | 43    | 138,680  | 0.9920      | 0.9930      | IRS  |
| MZ  | 1286  | 18.251     | 525    | 44    | 195,519  | 0.9860      | 1           | DRS  |
| OL  | 2258  | 20.646     | 734    | 54    | 179,742  | 0.7890      | 0.8160      | DRS  |
| OS  | 2170  | 15.335     | 649    | 47    | 171,159  | 0.9400      | 0.9870      | DRS  |
| PF  | 2375  | 10.613     | 422    | 27    | 128,899  | 0.9330      | 0.9600      | DRS  |
| R   | 1688  | 15.315     | 605    | 33    | 164,773  | 0.9010      | 0.9380      | DRS  |
| S   | 5236  | 81.157     | 809    | 64    | 491,316  | 0.5820      | 1           | DRS  |
| SB  | 2874  | 25.029     | 573    | 26    | 188,917  | 0.5910      | 0.6320      | DRS  |
| SN  | 2821  | 18.772     | 810    | 72    | 193,132  | 1           | 1           | CRS  |
| WI  | 2926  | 29.215     | 918    | 72    | 340,817  | 0.9590      | 1           | DRS  |
| WU  | 2120  | 13.481     | 429    | 30    | 129,125  | 0.7450      | 0.7680      | DRS  |
We then choose a special member of the group, namely the theatre of the town of Oldenburg and submit it to the process explained in foregoing sections. Theatre OL starts with 2258 seats, 20.646 mio. personnel expenses, and realizes 734, 54, and 179,742 units of the three outputs. Its relative CCR efficiency is 0.789 and its BCC efficiency amounts to 0.816. For the optimal price system, we have $\eta_{KL}^\pm = -0.07$, thus plighting under-proportional loss of outputs when reducing inputs. Furthermore, we presume an input reduction factor $f_{OL} = 0.98$. With these parameters, Oldenburg theatre from planning period to planning period would run through input/output sceneries, and CCR-, BCC-, and scale efficiencies (SE) like in Table 4 (rounded values).

From this table, we learn that the reduction process of activities for Oldenburg’s public would be very painful. First, reducing activities stepwise by $r^+ = 1.0703, 1.0773, 1.0856$, and 1.0956 plus input reduction of $(1 - 0.98) \times 100 = 2\%$ in each step means an irresponsible sellout of cultural quality in town. And such a sellout very likely will cause an angry protest in Oldenburg’s population. Even worse, this reduction does not make OL theatre efficient at all. All together, this strenuous effort over 4 planning periods results in 88.45% CCR and BCC efficiency and a 100% scale efficiency. However, mpss is still far away....

Data envelopment analysis is a suitable instrument for a DMU to detect its weaknesses and its improvement potentials. Whether or not a DMU can realize such findings depends on the surrounding conditions and the DMU’s change management, however. Whether or not a theatre like the one in Oldenburg will follow recommendations to reduce activities is by no means a DEA question but rather a political issue. We hope that the audience in this town will have many years of vivid sensations with its theatre.

**Conclusion and the road ahead**

In this contribution, radial returns to scale are measured for all DMUs along their respective (in-)efficiency hyper-planes. This measure involves not only the variable $u$ but also outputs and their corresponding virtual prices. The measure is valid for efficient and inefficient activities and can be applied even to cases with non-unique $u$’s. In other words: even for an interior point of the technology, unique and non-unique returns are measurable without any projection upon the technology boundary. This measure for each DMU is a handy instrument to evaluate consequences of radial upsizing and downsizing. Each time, such upsizing or downsizing is realized, the DMU might again check its efficiency and its returns to scale and take action to improve its situation. In this paper, we propose an algorithm to support DMUs in finding an economically reliable path towards mpss. This algorithm interactively communicates with the decision-maker to avoid unrealistic steps of activity improvements. In an application, such steps might be modified due to environmental conditions. These modifications are a worthwhile focus for future research.

Cross-efficiencies are considered an interesting approach to evaluate DMUs’ efficiencies from the point of view of other DMUs. Whenever a supervising institution dismisses self-appraisal as a valid concept, crosswise evaluations might help to find a peer, a weight system acceptable for all DMUs. Earlier and recent DEA literature reports on such peer-appraisal concepts confer, e.g., Doyle and Green (1994), Rödder and Reucher (2012). Once such a concept is accepted and a peer is selected, his weight system not only appraises all DMUs efficiencies but becomes the transfer price system of the whole group. Do there exist cross-RTS similar to cross-efficiencies and which consequences do such cross-RTS have upon a DMUs scale-sizing? Such questions could be the issue of further research.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**References**

Banker RD (1984) Estimating most productive scale size using data envelopment analysis. Eur J Oper Res 17:35–44

Banker RD, Thrall RM (1992) Estimation of returns to scale using data envelopment analysis. Eur J Oper Res 62:74–84

---

**Table 4** Interactive iterations for DMU OL

| Iteration | $\langle x_{OL}, y_{OL} \rangle$ | $g_{OL}^\ast$ | $g_{OL}^\ast$ | SE | $r^+$ | $f_{OL}$ |
|-----------|---------------------------------|----------------|----------------|----|-------|---------|
| OL        | (2258, 20.646, 734, 54, 179742)   | 0.7890         | 0.8160         | 0.9669 | -     | -       |
| 1         | (2055, 18.789, 686, 50, 167930)   | 0.8104         | 0.8325         | 0.9735 | 1.0703 | 0.98    |
| 2         | (1856, 16.970, 637, 47, 155880)   | 0.8329         | 0.8494         | 0.9806 | 1.0773 | 0.98    |
| 3         | (1661, 15.187, 586, 43, 143600)   | 0.8573         | 0.8686         | 0.9890 | 1.0856 | 0.98    |
| 4         | (1470, 13.442, 535, 39, 131070)   | 0.8845         | 0.8845         | 1     | 1.0956 | 0.98    |
Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30:1078–1091

Bashiri M, Farshaf-Geranmayeh A, Mogouie H (2013) An alternative transformation in ranking using l1-norm in data envelopment analysis. J Ind Eng Int 10(3):1–10

Charnes A, Cooper WW (1984) The non-Archimedean CCR ratio for efficiency analysis: a rejoinder to Boyd and Färe. Eur J Oper Res 15:333–334

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Coelli TJ, Prasada Rao DS, O’Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, vol 2. Springer, New York

Debreu G (1951) The coefficient of resource utilization. Econometrica 19:273–292

Dellnitz A (2016) RTS-mavericks in data envelopment analysis. Oper Res Lett 44(5):622–624

Deutscher Bühnenverein (2013) Theaterstatistik 2011/2012. Deutscher Bühnenverein Bundesverband der Theater und Orchester, Köln

Doyle J, Green R (1994) Efficiency and cross-efficiency in DEA: derivations, meanings and uses. J Oper Res Soc 45:567–578

Du J, Liang L (2012) Centralized production planning based on data envelopment analysis. Asia Pac Manag Rev 17:211–232

Emrouznejad A, Yang G (2017) A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socio Econ Plan Sci. doi:10.1016/j.seps.2017.01.008

Farell MJ (1957) The measurement of productive efficiency. J R Stat Soc 120:253–290

Forsund FR (1996) On the calculation of the scale elasticity in DEA models. J Product Anal 7:283–302

Forsund FR, Hjalmarsson L (2004) Calculating scale elasticity in DEA models. J Oper Res Soc 55:1023–1038

Forsund FR, Hjalmarsson L, Krivonozhko VE, Utkin OB (2007) Calculation of scale elasticity in DEA models: direct and indirect approaches. J Product Anal 28:45–56

Frisch R (1965) Theory of production. Reidel, Dordrecht

Fukuyama H (2000) Returns to scale and scale elasticity in data envelopment analysis. Eur J Oper Res 125:93–112

Golany B, Yu G (1997) Estimating returns to scale in DEA. Eur J Oper Res 103:28–37

Homayounfar M, Amirtreimooiri AR, Toloie-Eshlaghy A (2014) Production planning considering undesirable outputs—a DEA based approach. Int J Appl Oper Res 4:1–11