A comment on multi-stage DEA methodology

Laurens Cherchye *, Tom Van Puyenbroeck

Center for Economic Studies, Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

Received 24 May 1999; accepted 20 June 2000

Abstract

Coelli (Oper. Res. Lett. 23 (1998) 143) introduced a multi-stage (MS-) DEA methodology that represents an interesting alternative for the treatment of slacks remaining after proportional correction in inputs or outputs. By solving a sequence of radial models, one thus arrives at the identification of "more representative efficient points" (in terms of similar input/output mixes). We show that "most representative efficient points" can be found using a direct approach and may differ from those obtained by MS-DEA. Nevertheless, MS-DEA remains very attractive in view of its economic (intrinsic price) legitimation.

Keywords: Multi-stage DEA; Mix deviation; Cost interpretation

1. Introduction

As a solution to the slack problem of radial DEA measures, Coelli [6] proposed a multi-stage (MS-) DEA methodology that helps to identify "more representative efficient points" for the orientated DEA models relative to those obtained by the original two-stage linear programming process presented in [1]. Specifically, he proposes to proceed in each further step by radially scaling down that (input/output) subvector in which the evaluated observation is dominated in the Koopmans [10] sense. His conclusion is that "this approach seeks to identify efficient projected points which have input and output mixes which are as similar as possible to those of the inefficient points" (p. 148). However, he gives no precise specification of how mix deviation should be measured. In this paper we show that, for an intuitive definition of the concept, the above assertion is not strictly correct. Nevertheless, because of its computational convenience, MS-DEA remains recommendable. Moreover, it can be given an attractive economic legitimation as we subsequently demonstrate. Some remarks are contained in the final section.

We use $\leq$ ($\geq$) to indicate "less (greater) than or equal to", while $\leq$ ($\geq$) means $\leq$ ($\geq$) and $\neq$. Further, $\cos(u, z)$ represents $\cos \theta$ with $\theta$ being the angle between vectors $u$ and $z$. For ease of exposition, we focus solely on input efficiency in the following and discuss mix deviation in terms of input combinations. Extensions to output projections are straightforward, however. We further restrict attention to the constant returns to scale (CRS) DEA model presented...
2. Minimizing mix deviation

An appropriate indicator for the deviation between reference and evaluated input–output mixes appears to be the angle between the two vectors (see also [4]). In this respect, maximizing the cosine of this angle implies minimization of dispersion between the mixes. Obviously, standard DEA (radial) measures imply a cosine of value 1 as mixes of evaluated and reference input and output mixes, then implies maximizing the cosine of this angle to be the angle between the two vectors (see also [4]).

The maximum cosine is then associated with the most “resembling” reference in terms of closest input mix. Note that it may sometimes be computationally more efficient to proceed by identifying all efficient facets of the whole production possibility set (and not of each separate $L^{CRS}(y_i)$ individually) following, e.g. the algorithm of Yu et al. [15]. A straightforwardly analogous procedure then applies.

The “mix dispersion minimizing” reference will not necessarily coincide with that obtained from the MS-DEA algorithm as can be illustrated by means of a simple counterexample. Suppose we have three observations all of which produce the same amount of output. The input vectors are $x_1 = (10, 9, 15)$, $x_2 = (10, 7, 2)$ and $x_3 = (10, 5, 2, 8)$. The one-step radial procedure labels all three observations efficient. However, it is clear that only $x_2$ and $x_3$ are Koopmans efficient as $x_1$ has a slack in both the second and third inputs. Convex combinations of $x_2$ and $x_3$ constitute the efficient subset. The MS-DEA method will proceed in the second step by comparing collinearly the subvector $(9, 15)$ of $x_1$ to convex combinations of $(7, 2, 2)$ and $(5, 2, 8)$, which is illustrated in Fig. 1. The eventual MS-DEA projection for $x_1$ is $(10, 3, 5)$. The cosine of the angle between this reference and $x_1$ is 0.866. We can now apply model (2.2).

The cosine values are depicted in Fig. 2 as a function of the intensity parameter $x$, which gives the weight of $x_3$ in the reference vector. In Fig. 2 we also indicate the point $x = 0.5$, i.e. the value associated with the MS-DEA projection. This figure reveals that there are other combinations of $x_2$ and $x_3$ with a higher cosine.
value. It appears that mix deviation is minimized for \( x_3 \) and is the most representative reference. The associated cosine value between \( x_1 = (10, 9, 15) \) and the evaluated input vector \( (\sum_{j=1}^{P} x_j \hat{x}_j) / \sum_{j=1}^{K} \hat{x}_j \) is 0.922, which is clearly above 0.866. A point that deserves special attention is the fact that the cosine measure (2.2) is not invariant to the input vector \( e \) for all \( l, l' \in \{1, \ldots, K \} \) (i.e. the cosine maximizing projection is the radial projection) then the cosine will equal unity as the shrinkage vector is collinear to the unit vector. On the other hand, the more \( \lambda(x) \) deviates from the unit vector \( e \) in “mix” terms, the lower is the cosine value. This is intuitively appealing as in that case the proportional shrinkage factors (i.e. the \( \hat{x}_j(x) \)) are increasingly unequal. It is easy to check that applying this prescaling procedure to our example leads to exactly the same qualitative results, i.e. \( x_3 \) would be obtained as the “most representative reference” for \( x_1 \).

An interesting by-product of this last procedure is that it allows for immediately constructing an efficiency measure as a useful complement to the projection procedure, viz., by aggregating the obtained (cosine maximizing) references. For example, one could simply take the arithmetic mean (i.e. \( \sum_{l=1}^{K} \lambda_l(x) / K \)) or the harmonic mean (i.e. \( K / \sum_{l=1}^{K} (\lambda_l(x))^{-1} \)) as an efficiency gauge (compare with [8, 5], respectively, although we emphasize that the obtained efficiency measure need not be associated with the same projection point as identified in [8, 5]). We have thus established that MS-DEA does not necessarily yield most representative Koopmans efficient references in terms of mix properties. One could claim that this finding depends on how mix deviation is measured, viz., by means of the angle between the

\[
\max_{\lambda} \frac{1}{\sqrt{K}} \frac{\sum_{l=1}^{K} \lambda_l(x)}{\sqrt{\sum_{l=1}^{K} (\lambda_l(x))^2}}
\]
two vectors. However, the latter appears to be a natural choice and it is unlikely that our result would be contradicted for any other indicator.

While the above suggests that Coelli’s original motivation for MS-DEA is debatable, the projection procedure clearly has very attractive features. As pointed out by Coelli [6], it is units invariant (no prescaling is hence required) and computationally not very demanding. Last but not least, there exists an economic justification for MS-DEA as we will show in the next section.

3. An economic legitimation for MS-DEA

Let us denote the radial (input) efficiency estimate associated with an input–output vector \((x_i, y_i)\) by \(\rho_i\). Debreu [7] proved that this measure can also be expressed as follows:

\[
\rho_i \equiv \max_{x} \frac{(p_R)^T x_R}{(p_R)^T x_i}, \quad \text{with } x_R \in \text{Isoq} \, L^{\text{CRS}}(y_i),
\]

(3.1)

where \(p_R \geq 0\) is a shadow (or “implicit”) price vector associated with \(x_R\) (i.e. \((p_R)^T x_R \leq (p_R)^T x_i \forall x_i \in L^{\text{CRS}}(y_i)\)) and \(\text{Isoq} \, L^{\text{CRS}}(y_i)\) the isocost of \(L^{\text{CRS}}(y_i)\):

\[
\text{Isoq} \, L^{\text{CRS}}(y_i) \equiv \{x \mid x \in L^{\text{CRS}}(y_i), \lambda x \notin L^{\text{CRS}}(y_i) \text{ for } \lambda \in (0, 1)\}. \quad (3.2)
\]

Clearly, from comparison of (2.1) with (3.2), \(\text{Eff} \, L^{\text{CRS}}(y_i) \subseteq \text{Isoq} \, L^{\text{CRS}}(y_i)\): radially efficient input combinations are not necessarily Koopmans efficient. So, (3.1) reveals that radial projection results in a reference belonging to the isocost that maximizes the ratio of reference to actual costs (referred to as “cost ratio” in the following) evaluated against a shadow price vector associated with the latter (this shadow price vector need not necessarily be unique and is determined up to a positive multiplicative scalar). This gives a nice economic legitimation for (one-step) radially projecting on the production frontier. That is, the radial measure has an evocative characterization as an upper bound to economic efficiency using the shadow prices implicit in the production technology (see also [12]).

The slack problem of radial DEA measures boils down to the possibility of zero elements in \(p_R\). In the example of the previous section a zero implicit price is associated with the second and third inputs of \(x_1\). On the other hand, Koopmans [10] showed that in a linear activity setting (like DEA) each element of the efficient subset corresponds to at least one strictly positive shadow price vector. This Koopmans characterization of efficient production is widely accepted because of its intimate link with the first two fundamental theorems of welfare economics, which say that (for this specific setting) a productive organization is (input and output) efficient if and only if one can construct a strictly positive price vector under which it becomes profit maximizing. Analogously, in the input space a vector is efficient if and only if there exists a strictly positive price combination under which it is cost minimizing. In our example this is the case for \(x_2\), \(x_3\) and all their convex combinations.

The MS-DEA projection obviously meets the Koopmans definition. Moreover, it can be shown to maximize the cost ratio, but now precisely evaluated against a strictly positive shadow price vector. To see this, suppose that the one-step radial procedure projects \(x_i\) on \(x_R \in \text{Isoq} \, L^{\text{CRS}}(y_i) \setminus \text{Eff} \, L^{\text{CRS}}(y_i) (x_R = p_R x_i)\). Denote the set of all \(x_E \in \text{Eff} \, L^{\text{CRS}}(y_i)\) with the same amount of inputs in which \(x_R\) does not exhibit slack by \(X_E \equiv \{x \mid x \in \text{Eff} \, L^{\text{CRS}}(y_i) \text{ and } x \leq x_R\}\). Clearly, all elements of \(X_E\) have the same cost ratio as \(x_R\) (with zero prices accorded to those dimensions in which there is slack). E.g., in the above example this applies for convex combinations of \(x_2\) and \(x_3\) when \(x_1\) is evaluated.

For all other \(x'_E \in \text{Eff} \, L(y_i) \setminus X_E\) the cost ratio is strictly exceeded by that of \(x_R\). Each element of \(X_E\) is by construction also associated with a strictly positive shadow price vector. Thus, by choosing the shadow prices corresponding to those dimensions in which \(x_R\) has a slack small enough compared to those of the non-slack dimensions, we still have that the resulting cost ratio related to each \(x_E \in X_E\) will strictly be above the cost ratios of all \(x'_E \in \text{Eff} \, L(y_i) \setminus X_E\). That is, we are led to choose among the \(x_E \in X_E\) when searching for (in terms of cost ratios) a suitable Koopmans efficient reference.

Split up \(x_{R1}\) in a slack subvector \(x_{RS}\) and a non-slack subvector \(x_{RN}\), and similarly decompose the shadow price vector \(p_R\) associated with \(x_R\) in \(p_{RS} (= 0)\) and \(p_{RN} (> 0)\). In the second step only \(x_{RS}\) is further
reduced. Also decompose $x_i$ and $x_E (\in X_E)$ in $x_S$ and $x_N$ (so that $x_{iN} = \rho_i x_N$ ($i = N,S$) and $x_{iS} \neq \rho_i x_N$), respectively. One thus has to find $x^*$ (decomposable in $x_R^* \leq x_S$ with associated implicit price vector $p_R^* \geq 0$ and $x_{RS} = \rho_i x_N$ with, without loss of generality, $p_R^* = p_{RS}$) such that:

$$
\frac{\rho_i(p_{RS})^T x_{iN} + (p_{RS})^T x_{iS}}{(p_{RS})^T x_S + (p_{RS})^T x_{iS}} \geq \frac{\rho_i(p_{RS})^T x_{iN} + (p_{ES})^T x_{iS}}{(p_{ES})^T x_S + (p_{ES})^T x_{iS}},
$$

(3.3)

where $x_{ES}$ corresponds to the $x_E \in X_E$ with a shadow price vector $p_{ES} \geq 0$. (3.3) can be formulated as

$$
\frac{\rho_i + (p_{RS})^T x_{iS}}{1 + (p_{RS})^T x_{iS}} \geq \frac{\rho_i + (p_{ES})^T x_{iS}}{1 + (p_{ES})^T x_{iS}},
$$

(3.4)

where $(p_{RS})^T x_{iS} = (p_{ES})^T x_{ES} = c$, with $c$ being a positive constant. To see this, note that (3.3) should hold for all rescalings of the vectors $p_{RN}$ and $p_{PS}$ ($k = R^*, E$). Next, we use that $x_{RS}, x_{iS}$ and $x_S$ are radially efficient, i.e. $(p_{RN})^T x_{iS} = (p_{RN})^T x_{iN}$ and $(p_{RN})^T (\rho_i x_N) + (p_{PS})^T x_{iS} \leq (p_{RN})^T x_{iS} + (p_{PS})^T x_{iS} (k = R^*, E)$ for all $x_i = (x_{iN}, x_{iS}) \in L^*$. Convex combinations of these conditions (with $\kappa_i \in [0,1], k = R^*, E$) yield $(p_{RN})^T x_{iS} = \kappa_i (p_{ES})^T x_{iS}$, from which we obtain $(p_{RS})^T x_{iS} = (p_{ES})^T x_{ES} \equiv c$ for $(p_{ES})^T x_{iS} = [(\kappa_i (p_{ES})^T x_{ES})]/(p_{ES})^T (\rho_i x_N))$ (4.1). Condition (3.4) becomes

$$
(\rho_i + c)((p_{ES})^T x_{iS} - (p_{ES})^T x_{iS}) \geq 0.
$$

(3.5)

Eq. (3.5) is fulfilled when $x_{RS}$ is obtained from equiproporionate reduction of $x_{RS}$, as from (3.1):

$$
\frac{(p_{RS})^T x_{RS}}{\rho_i(p_{RS})^T x_{iS}} \geq \frac{(p_{ES})^T x_{ES}}{\rho_i(p_{ES})^T x_{iS}}
$$

or $(p_{ES})^T x_{iS} - (p_{RS})^T x_{iS} \geq 0$.

This establishes an economic legitimation for applying radial (subvector) projection in the second step. Of course, it could be that $x_{RS} \not\in X_E$ (some elements of $p_{RS}$ are equal to zero). In this case one can formulate a straightforwardly analogous argument to justify a radial procedure in the third and (if necessary) each further step.

Summarizing, the MS-DEA algorithm has a sound economic basis which is very similar to the one-step procedure, but strengthens the latter by making shadow prices in the cost ratio strictly positive. There always exists a pair of strictly positive implicit price vectors under which the MS-DEA cost ratio exceeds that associated with any other element of the efficient subset.

4. Conclusion

When gauging the relative performance of a productive unit it often seems defendable to use benchmarks that are its peers in the most genuine sense. This is after all one intuitive rephrasing of the benefit-of-the-doubt perspective underlying a great deal of the (in)efficiency measurement literature. On the other hand, sticking to Farrell’s [9] “most conservative assumption” regarding mixes can yield peers that are themselves relatively inefficient. Identifying a mix-minimizing efficient reference in the way described above safeguards the conservative assumption to the furthest extent possible. In particular, it will weakly dominate MS-DEA in identifying the ‘best’ efficient peers. Evidently, this yields a “most representative efficient reference” only to the extent that one indeed interprets the general notion of representativeness in terms of product mixes. Or stated otherwise, the produced peers can still be criticized in view of the fact that an (almost) equiproportionate rescaling of input/output vectors is not always implementable in practice.

Although the latter issue may occasionally arise in evaluation practice, the basic idea of radial correction is still defended by many in view of its intuitive economic interpretation as a benefit-of-the-doubt cost ratio. Adhering to this framework, the wasteful production associated with slacks can be interpreted as the presence of implicit zero prices in that ratio. When slacks are removed using MS-DEA this effectively implies that the economic interpretation can be strengthened to a benefit-of-the-doubt cost ratio with strictly positive implicit prices. Even if it were only on
this account, MS-DEA seems a particularly appealing method for the treatment of slacks in the programming approach to productive efficiency measurement.

Acknowledgements

We want to thank Tim Coelli, Holger Scheel and an anonymous referee for helpful comments on previous versions of this paper.

References