Journal of the Operational Research Society (2011) 62, 1413–1419

www.palgrave-journals.com/jors/

Negative data in DEA: a simple proportional distance function approach

K Kerstens[1]* and I Van de Woestyne[2]

[1]IESEG School of Management, Lille, France; and [2]Hogeschool Universiteit Brussel, Brussels, Belgium

The need to adapt Data Envelopment Analysis (DEA) and other frontier models in the context of negative data has been a rather neglected issue in the literature. A recent article in this journal proposed a variation on the directional distance function, a very general distance function that is dual to the profit function, to accommodate the occurrence of negative data. In this contribution, we define and recommend a generalised Farrell proportional distance function that can do the same job and that maintains a proportional interpretation under mild conditions.

Journal of the Operational Research Society (2011) **62,** 1413–1419. doi:10.1057/jors.2010.108
Published online 28 July 2010

Keywords: data envelopment analysis; linear programming; optimization; production; finance

1. Introduction

The seminal article of Farrell (1957) and the revived interest of Charnes *et al* (1978) have led to the development of the Data Envelopment Analysis (DEA) literature that has developed at the interface of operational research and economics (see, eg, Førsund and Sarafoglou (2005) for its history). This DEA literature has meanwhile become one of the success stories of the operational research area (see, eg, Emrouznejad *et al*, 2008). The estimation of frontier or best practice models to determine the relative efficiency of organisational units has found its way into a large variety of domains of application. In terms of empirical surveys of certain well-analysed sectors, one could, for instance, point to banking (eg, Harker and Zenios, 2001), education (Worthington, 2001), health care (eg, Ozcan, 2008), insurance (Cummins and Weiss, 2000), public transit (eg, De Borger *et al*, 2002), and real estate (Anderson *et al*, 2000). In addition to this surge of empirical applications, there has been an extended series of methodological developments in this literature (see, eg, the surveys in Färe *et al* (1994) or Thanassoulis *et al* (2008)).

In a traditional production context, inputs and outputs are assumed to be non-negative (see, eg, Färe *et al* (1994) for conditions on the input and output data matrices). However, frontier applications have also moved into areas where negative data may occur. Examples include the analysis of financial statements (eg, Smith (1990) or Feroz *et al* (2003)) or the rating of mutual funds (see the seminal article by Murthi *et al*, 1997), etc. Obviously, growth rates or returns can be both negative and positive (see Sharp *et al* (2007) for a discussion of contexts where negative inputs and outputs arise naturally).

The issue of handling negative data has attracted some research attention. For instance, proposals have been made to translate the data (eg, by adding a number making all data positive), though in many models this may have implications on the efficiency measures, among other things (see, eg, Ali and Seiford, 1990). In fact, very few DEA models turn out to yield solutions that are invariant to such data transformations (ie, are translation invariant). A number of other solutions have been proposed in the DEA literature (eg, Silva Portela *et al* (2004), Sharp *et al* (2007), among others). This small literature has been competently summarised in Pastor and Ruiz (2007) or Thanassoulis *et al* (2008).

The rather recently introduced directional distance function generalises existing distance functions by accounting for both input contractions and output improvements, and it is dual to the profit function (see Chambers *et al*, 1998). Luenberger (1992) introduced the benefit function as a directional representation of preferences generalising the input distance function defined in terms of the utility function. Luenberger (1995) transposed this benefit function in a production context under the name of the shortage function. Chambers *et al* (1998) relabel this same function as a directional distance function and this name has become its most common denomination. This directional distance function is flexible due to the variety of direction vectors it allows for. In the more pragmatic, managerially oriented benchmarking models allowing for negative data, Silva Portela *et al* (2004) suggest working with some variations of this directional distance function.

*Correspondence: K Kerstens, CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, Lille F-59000, France.
E-mail: k.kerstens@ieseg.fr

In this contribution, we argue that a very simple modification of the traditionally defined proportional distance function can equally well be used to accommodate for negative data.

2. Technology and directional distance function

The standard function of a production technology is to transform inputs $x=(x_1, \ldots, x_p)\in\mathbb{R}^p_+$ into outputs $y=(y_1, \ldots, y_q)\in\mathbb{R}^q_+$. The production possibility set or technology T summarises the set of all feasible input and output vectors and can be defined as follows:

$$T = \{(x, y) \in \mathbb{R}^{p+q}_+; x \text{ can produce } y\}. \tag{1}$$

The technology satisfies the following standard assumptions: (T.1) no free lunch; (T.2) boundedness; (T.3) closedness; (T.4) strong disposal of inputs and outputs; and (T.5) convexity (see Färe *et al* (1994) for details). For a finite data set, it has been demonstrated that the DEA model suggested by Banker *et al* (1984) is the non-parametric minimum extrapolation technology satisfying a subset of the above set of axioms (see also Färe *et al*, 1994). Throughout this contribution, following the literature, we refer to this model as a convex, strongly disposable technology, satisfying variable returns to scale (VRS).

The technology can be characterised by the use of distance functions. To simplify notation, denote the netput vector $z=(x,y)\in T$ and the direction vector $g=(h,k)\in(-\mathbb{R}^p_+)\times\mathbb{R}^q_+$, that is partitioned in an input and an output direction vector $-h$ and k, respectively. The directional distance function is seeking a simultaneous improvement in both the input and output dimensions in the direction of the vector g and is formally defined as:

Definition 2.1 For a given technology T, the directional distance function D_T is the function D_T: $T\times((-\mathbb{R}^p_+)\times\mathbb{R}^q_+)\to\mathbb{R}\cup\{+\infty\}$ with

$$D_T(z; g) = \sup_{\delta}\{\delta \in \mathbb{R} : z + \delta g \in T\}.$$

First, observe that, by extending the target set $\mathbb{R}$ with $+\infty$, the directional distance function is well defined for all possible choices of the direction vector. Indeed if $g=0$, then clearly $D_T(z;0)=+\infty$. Also note that $D_T(z;g)\geqslant 0$, since $\delta=0$ is always contained in the set $\{\delta\in\mathbb{R}: z+\delta g\in T\}$. Note that in the more general case where a point may not be part of the technology, the definition of the directional distance function must be adapted such that it distinguishes between the standard case where the supremum occurring in the definition is finite and the case where it is infinite ($-\infty$ or $+\infty$) (the latter showing up as infeasibilities in the associated mathematical programming problems). This distinction is important since Briec and Kerstens (2009) have recently shown for general production technologies that there are always circumstances for which this adapted function may not be well defined. For instance, this leads to a lack of determinateness in the Luenberger productivity indicator (a generalisation of the more widely applied Malmquist productivity index) that is defined using a combination of directional distance functions: in particular, infeasibilities may occur when relating an adjacent period observation to a technology in a given time period (see Briec and Kerstens (2009) for details).

Second, this distance function has an interpretation as an efficiency (or better, inefficiency) measure, because it measures deviations from the boundary of the technology. A weakly efficient vector $z\in T$ yields a directional distance function value of zero.

The directional distance function has proven to be a useful tool in applied production analysis. For instance, it allows Chavas and Kim (2007) to shed new light on economies of scope from a primal viewpoint and to propose a decomposition of the benefits from integrated *versus* (partially or completely) specialised firms that includes output complementarities, economies of scale, and convexity. Furthermore, it provides the defining components of the Luenberger productivity indicator (eg, Chambers, 2002), a generalisation of the very popular Malmquist productivity index. While the latter index is based on input- or output-oriented distance functions (that are inversely related to the traditional radial input- or output-oriented efficiency measures), the former indicator is based on directional distance functions and is therefore suitable to evaluate total factor productivity growth for profit-oriented institutions. Both productivity indices can be decomposed into a technical efficiency change and a frontier change component, which partly explains their popularity.

We mention the following proposition that can be obtained directly from Definition 2.1.:

Proposition 2.1 *For a given technology* T, $z\in T$, $g\in(-\mathbb{R}^p_+)\times\mathbb{R}^q_+$ *and an arbitrary norm function* $\|\ldots\|$, *it follows that* $D_T(z;g)=\delta^*=\|z^*-z\|/\|g\|$, *with* $z^*=z+\delta^* g$.

Proof The proof is trivial and is therefore omitted. □

The directional distance function defined in Definition 2.1 uses a general direction vector g. However, sometimes one considers the special case $\bar{g}=(-x,y)$ which gives rise to the Farrell proportional distance function (Briec, 1997) described in the following definition:

Definition 2.2 For a given technology T, the Farrell proportional distance function F_T is the function

$$F_T : T \to \mathbb{R}\cup\{+\infty\} : z \mapsto F_T(z) = D_T(z; \bar{g}),$$

with $z=(x,y)$ and $\bar{g}=(-x,y)$.

Since this proportional distance function is a special case of the directional distance function, it also measures inefficiency. Note that with this choice, $\bar{g} \in (-\mathbb{R}^p_+) \times \mathbb{R}^q_+$. Obviously, given semi-positive prices, looking for reductions in inputs and expansions in outputs contributes to the objective of profit maximisation.

The axiomatic foundations of efficiency measures in production theory have been analysed since at least Färe and Lovell (1978). This seminal article proposed three axioms that an input-based efficiency index should satisfy: (i) indication (ie, index equals unity if and only if the input vector belongs to the strongly efficient subset), (ii) monotonicity (ie, for constant other inputs and outputs, increasing an input must reduce the value of the index), and (iii) homogeneity of degree minus one (ie, doubling inputs must halve the index). Additional axioms have been proposed (eg, commensurability, continuity in technology and in input or output quantities, etc). This literature focused mainly on special distance functions that only look for a reduction in inputs (or improvements in outputs). However, the directional distance function measures potential efficiency improvements in all dimensions. Axiomatic properties of the directional distance function and the Farrell proportional distance function are studied in Chambers *et al* (1998) and Briec (1997), respectively. Russell and Schworm (2009) recently took a look at similar efficiency measures in production theory and prudently conclude that the directional distance function with a proportional interpretation satisfies a stronger unit invariance property compared to the case of a fixed direction. However, this axiomatic literature is not central to our contribution. Therefore, we focus on the proportional interpretation one can attribute to certain variations of the directional distance function.

We first introduce the following definition:

Definition 2.3 A norm function $\| \ldots \|$ on $\mathbb{R}^n$ is reflection invariant if $\|z\| = \|f_s(z)\|$ for all $z \in \mathbb{R}^n$ and all reflections f_s with respect to an arbitrary hyperplane $S \subset \mathbb{R}^n$.

It can be easily shown that the Euclidean norm has the property of being reflection invariant.

The meaning of proportionality of the Farrell proportional distance function is twofold as can be seen from the following proposition:

Proposition 2.2 *For a given technology T, $z \in T$ and some reflection invariant norm function $\| \ldots \|$, the Farrell proportional distance function value $F_T(z)$ satisfies:*

(a) $F_T(z) = \bar{\delta}^* = \dfrac{\|z^* - z\|}{\|\bar{g}\|} = \dfrac{\|z^* - z\|}{\|z\|}$,
with $z^* = z + \bar{\delta}^* \bar{g}$;

(b) $0 \leqslant F_T(z) \leqslant 1$.

Proof Both statements follow from Proposition 2.1. (a) Indeed, since a reflection invariant norm function is assumed, $\|\bar{g}\| = \|(-x, y)\| = \|(x, y)\| = \|z\|$. When combined with Proposition 2.1, (a) follows directly. (b) Let $z^* = (x^*, y^*)$. Since $z^* = z + \bar{\delta}^* \bar{g}$, it is also the case that $x^* = x + \bar{\delta}^*(-x) = (1 - \bar{\delta}^*)x$. However, both x and x^* are contained in $\mathbb{R}^p_+$, with results in $\bar{\delta}^* \leqslant 1$. Obviously, since norm functions are positive, $\bar{\delta}^* \geqslant 0$, which concludes (b). □

Obviously, a percentage interpretation facilitates the utilisation of benchmarking results by practitioners.

Now, consider n decision-making units (DMUs) $z_i = (x_i, y_i)$, $(i = 1, \ldots, n)$ from which the technology T is derived. Furthermore, $z_0 = (x_0, y_0)$ denotes the DMU under analysis and $g = (h, k)$ is the selected direction vector. Then, the directional distance function value $D_T(z_0; g)$ under convexity, VRS, and strong disposability assumptions is obtained by solving the following linear programming (LP) problem:

$$D_T(z_0; g) = \max \left\{ \delta : \sum_{i=1}^{n} \lambda_i x_{ir} \leqslant x_{0r} + \delta h_r, \quad (r = 1, \ldots, p), \right. \\ \sum_{i=1}^{n} \lambda_i y_{is} \geqslant y_{0s} + \delta k_s, \quad (s = 1, \ldots, q), \\ \left. \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geqslant 0, \quad (i = 1, \ldots, n) \right\}. \quad (2)$$

From (2), it is clear that the Farrell proportional distance function value for the same technology can be computed by solving the following LP problem:

$$F_T(z_0) = \max \left\{ \bar{\delta} : \sum_{i=1}^{n} \lambda_i x_{ir} \leqslant x_{0r} - \bar{\delta} x_{0r}, \quad (r = 1, \ldots, p), \right. \\ \sum_{i=1}^{n} \lambda_i y_{is} \geqslant y_{0s} + \bar{\delta} y_{0s}, \quad (s = 1, \ldots, q), \\ \left. \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geqslant 0, \quad (i = 1, \ldots, n) \right\}. \quad (3)$$

3. Proportional distance function: a reformulation for negative data

Assuming now that inputs and/or outputs can be negative, one must revise the notion of a technology. In fact, an element of T no longer needs to be contained in $\mathbb{R}^{p+q}_+$. Hence, we redefine the technology on this extended data domain as

$$T' = \{(x, y) \in \mathbb{R}^{p+q}; x \text{ can produce } y\}, \quad (4)$$

with the standard assumptions stated before (except (T.1)).[1] Note that this extended data domain can yield problems for certain technology specifications (eg, the assumption of constant returns to scale: see Silva Portela *et al* (2004) for details). With this adaptation, Definition 2.1 of the directional distance function, the corresponding model (2) for computing it, and Proposition 2.1 remain all valid. However, the Farrell proportional distance function defined in Definition 2.2 is no longer well defined when inputs or outputs can take negative values, since the direction vector g is not necessarily contained in $(-\mathbb{R}^p_+) \times \mathbb{R}^q_+$. Such a choice is crucial to guarantee a simultaneous increase in the output direction and a decrease in the input direction. For instance, assume we start from an observation with one positive input and one negative output. Then, when applying model (3), one would be trying to obtain an even more negative output. Assuming an objective function is increasing in outputs and decreasing in inputs, this would lead to a decrease in the objective function value, which is undesirable.

To circumvent this problem, Silva Portela *et al* (2004) propose a so-called range directional model. In this model, the direction vector $\tilde{g} = (-R_0, S_0)$ is chosen for a DMU $z_0 = (x_0, y_0)$ with

$$R_{0r} = x_{0r} - \min\{x_{ir}; i = 1, \ldots, n\}, \quad (r = 1, \ldots, p);$$
$$S_{0s} = \max\{y_{is}; i = 1, \ldots, n\} - y_{0s}, \quad (s = 1, \ldots, q).$$

This choice guarantees a non-zero direction vector $\tilde{g} \in (-\mathbb{R}^p_+) \times \mathbb{R}^q_+$ under all circumstances, thereby realising the range directional distance function $R_{T'}$ suitable for negative as well as positive data. In the case of a technology-satisfying convexity, VRS, and strong disposability assumptions, the following LP problem needs to be solved:

$$R_{T'}(z) = \max \left\{ \tilde{\delta} : \sum_{i=1}^{n} \lambda_i x_{ir} \leqslant x_{0r} - \tilde{\delta} R_{0r}, \quad (r = 1, \ldots, p), \right.$$
$$\sum_{i=1}^{n} \lambda_i y_{is} \geqslant y_{0s} + \tilde{\delta} S_{0s}, \quad (s = 1, \ldots, q),$$
$$\left. \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geqslant 0, \quad (i = 1, \ldots, n) \right\}. \tag{5}$$

An obvious problem with this proposal is the fact that the efficiency measure resulting from the range directional model no longer has a proportional interpretation, which is a disadvantage for practitioners.[2]

However, there is another simple alternative that basically generalises the proportional distance function to handle negative data as well. This seems to have gone unnoticed in the literature so far. Given a DMU $z = (x_0, y_0)$, we propose the direction vector $\hat{g} = (-|x_0|, |y_0|)$ in which $|x_0|$ denotes the input vector with components $|x_{0r}|$, $(r = 0, \ldots, p)$, and similarly $|y_0|$ denotes the output vector with components $|y_{0s}|$, $(s = 0, \ldots, q)$ instead of taking $\bar{g}$ or $\tilde{g}$. Obviously, this choice assures that $\hat{g} \in (-\mathbb{R}^p_+) \times \mathbb{R}^q_+$ for both positive and/or negative data. Moreover, in the case of positive inputs and outputs, the direction vector coincides exactly with the one defining the original Farrell proportional distance function. Therefore, the proposed solution can indeed be seen as a generalisation of the Farrell proportional distance function suitable for both positive and negative data domains. We suggest calling it the generalised proportional distance function, which leads to the following definition:

Definition 3.1 For a given technology T', the generalised proportional distance function $G_{T'}$ is the function

$$G_{T'} : T' \to \mathbb{R} \cup \{+\infty\} : z \mapsto G_{T'}(z) = D_{T'}(z; \hat{g}),$$

with $z = (x, y)$ and $\hat{g} = (-|x|, |y|)$.

The proportional interpretation of the Farrell proportional distance function formulated in Proposition 2.2 now has a counterpart for this generalised proportional distance function:

Proposition 3.1 *For a given technology T', $z \in T'$ and some reflection invariant norm function $\| \ldots \|$, the generalised proportional distance function value $G_{T'}(z)$ satisfies*:

(a) $G_{T'}(z) = \hat{\delta}^* = \dfrac{\|z^* - z\|}{\|\hat{g}\|} = \dfrac{\|z^* - z\|}{\|z\|}$,
with $z^* = z + \hat{\delta}^* \hat{g}$;

(b) $0 \leqslant G_{T'}(z) \leqslant 1$ *if at least one of the input dimensions is strictly positive.*

Proof (a) Because the norm function is reflection invariant,

$$\|\hat{g}\| = \|(-|x|, |y|)\| = \|(|x|, |y|)\| = \|(x, y)\| = \|z\|.$$

Consequently, (a) follows from Proposition 2.1. (b) Let $z^* = (x^*, y^*)$. Since $z^* = z + \hat{\delta}^* \hat{g}$ it is also the case that $x^* = x + \hat{\delta}^*(-|x|)$. Assume the k-th input dimension is strictly positive. Then, $x^*_k = x_k + \hat{\delta}^*(-x_k) = (1 - \hat{\delta}^*) x_k$. Since both x_k and x^*_k are strictly positive for all data in the k-th dimension, $\hat{\delta}^* \leqslant 1$, with (b) as a consequence. □

Note that the ratio interpretation (statement (a) in Proposition 3.1) still holds for the generalised proportional distance function. However, the restriction of its function value between zero and one (statement (b) in Proposition 3.1) is now only guaranteed in the presence of

[1] In fact, it is often ignored that the model proposed by Banker *et al* (1984) does not satisfy the no free lunch axiom.

[2] This important contribution is further discussed and contrasted with other proposals regarding negative data in Pastor and Ruiz (2007).

at least one strictly positive input dimension within the sample. Note that in empirical applications the latter condition is rather mild: in the case of multiple input dimensions, it is rare to have a DMU with only negative values for all of these dimensions. For instance, for portfolio models using the shortage function (Briec and Kerstens, 2010) containing some even moment (eg, variance, kurtosis, ...) this condition is automatically met.

From model (3), it directly follows that the generalised proportional distance function value for a given DMU, under the same assumptions as above, is computed from the following LP model:

$$G_{T'}(z) = \max \left\{ \hat{\delta} : \sum_{i=1}^{n} \lambda_i x_{ir} \leqslant x_{0r} - \hat{\delta}|x_{0r}|, \quad (r = 1, \ldots, p), \right. \\ \sum_{i=1}^{n} \lambda_i y_{is} \geqslant y_{0s} + \hat{\delta}|y_{0s}|, \quad (s = 1, \ldots, q), \\ \left. \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geqslant 0, \quad (i = 1, \ldots, n) \right\}. \tag{6}$$

Note that the generalised proportional distance function value is a measure of inefficiency just like the proportional distance function. The closer this value is to zero, the more efficient the corresponding DMU.

Figure 1 illustrates the proposed direction vector on a theoretical example consisting of 65 DMUs with one input (X) and one output (Y). These DMUs are visualised by small grey circles. Both inputs and outputs can be negative. The DEA VRS frontier is determined completely by five DMUs defining the vertex points of this piecewise linear frontier. These vertices have as coordinates $(-12, -6)$, $(-9, 3)$, $(-4, 10)$, $(8, 15)$, and $(14, 17)$, respectively. For four DMUs (labelled with numbers 1–4), the projection onto the frontier by means of the generalised proportional distance function is indicated with an arrow, whereby the direction vector is selected to be $\hat{g} = (-|x_0|, |y_0|)$ for a given DMU $z = (x_0, y_0)$. Note that DMUs 1 and 2 (3 and 4) have a positive (negative) value for the single input.

Table 1 focuses on these four DMUs and their projections. The coordinates (x_0, y_0) of the DMUs labelled with numbers 1–4 are provided in columns 2 and 3. The coordinates of the direction vector $\hat{g} = (\hat{g}_x, \hat{g}_y)$ used in the generalised proportional distance function are listed in columns 4 and 5. Consequently, the directions of the arrows in Figure 1 are determined by the absolute value of the coordinates of the position vector of the initial points. Thus, despite what Figure 1 might suggest at first sight, the directions of the arrows are not arbitrary, but are precisely determined by the positions of the evaluated DMUs.

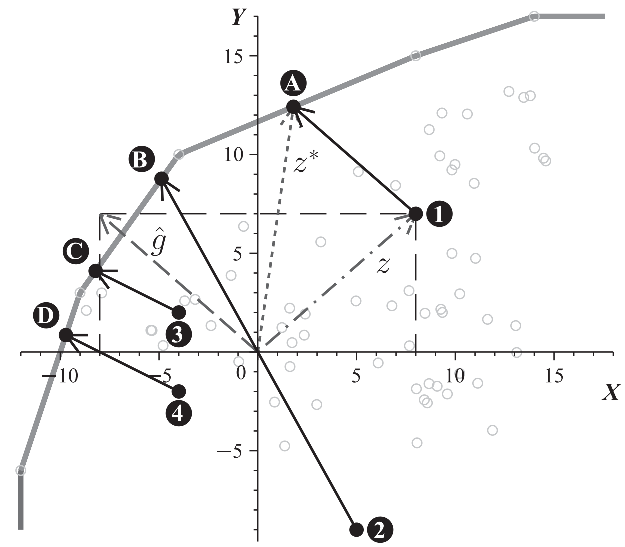

Figure 1 DEA VRS frontier: projections for four inefficient DMUs.

Table 1 Numerical example with four DMUs

From–To	x_0	y_0	$\hat{g}_x$	$\hat{g}_y$	x_0^*	y_0^*	d_x	d_y	$\hat{\delta}^*$
1–A	8	7	−8	7	1.806	12.419	−6.194	5.419	0.7742
2–B	5	−9	−5	9	−4.857	8.775	−9.857	17.775	1.9750
3–C	−4	2	−4	2	−8.211	4.105	−4.211	2.105	1.0526
4–D	−4	−2	−4	2	−9.714	0.857	−5.714	2.857	1.4286

In Figure 1, the resulting projection points located on the frontier are labelled with the letters *A–D*. Columns 6 and 7 in Table 1 represent the coordinates (x_0^*, y_0^*) of these projection points *A–D*. Projection points are computed by inserting the optimal value of the directional distance function in the right-hand side of the p and q inequalities related to the input and output dimensions, respectively. For example, point A on the frontier can be computed as follows: $(8-0.7742 \times 8, 7+0.7742 \times 7) = (1.806, 12.419)$.

The coordinates of the difference vector $d = (d_x, d_y) = (x_0^* - x_0, y_0^* - y_0)$ connecting the initial point with the projection point (visualised in Figure 1 with an arrow) is found in columns 8 and 9. Finally, the value of the generalised proportional distance function $\hat{\delta}^*$ for the four DMUs is found in the last column.

We point out that the optimal value of the generalised proportional distance function can easily be verified ex post from the previous elements in Table 1. We illustrate this for the DMU labelled 1. Consequently, the vector z corresponds with the dash-dotted arrow connecting the origin with the point labelled 1, and the vector z^* can be observed as the dotted arrow from the origin to the point labelled A. In addition, the corresponding direction vector $\hat{g}$, shown as the dashed arrow, is obtained by reflecting the dash-dotted arrow across the vertical axis. It follows from Proposition 3.1 that

$$G_{T'}(z) = \hat{\delta}^* = \frac{\|z^* - z\|}{\|\hat{g}\|} = \frac{\|A1\|}{\|\hat{g}\|} = \frac{\|A1\|}{\|z\|}, \qquad (7)$$

with $\|A1\| = \|z^* - z\|$ the distance from the point labelled 1 to the point labelled A. Note that using the notion of distance requires a reflection invariant norm function as indicated in Proposition 3.1. Therefore, we consider here the commonly used Euclidean norm for computing distances.

Consequently,

$$\begin{aligned}\hat{\delta}^* &= \frac{\|d\|}{\|\hat{g}\|} = \frac{\sqrt{d_x^2 + d_y^2}}{\sqrt{\hat{g}_x^2 + \hat{g}_y^2}} \\ &= \frac{\sqrt{(1.806 - 8)^2 + (12.419 - 7)^2}}{\sqrt{(-8)^2 + 7^2}} \\ &= \frac{\sqrt{(-6.194)^2 + 5.149^2}}{\sqrt{(-8)^2 + 7^2}} = 0.7742. \qquad (8)\end{aligned}$$

The inefficiency measures for the other points can be computed in a similar fashion.

We first recall that Proposition 3.1 guarantees a proportional interpretation of the inefficiency measure. In this example, however, its value can be larger than one since the only input present can attain negative values. This can be observed for the DMUs labelled 2, 3, and 4. From Proposition 3.1 statement (a), we know this means the (Euclidean) distance from the DMU to the optimal frontier is then larger than the (Euclidean) distance to the origin. For instance, for DMU 2, the efficiency measure amounts to 1.9750, or 197.50%. This means that the (Euclidean) distance of this DMU to the optimal frontier (location labelled B) is almost double the (Euclidean) distance to the origin. In Figure 1, this can be observed from the fact that the origin is located nearly halfway between the points labelled 2 and B. Obviously, the closer a point is situated to the frontier, the smaller is the numerator of (7) leading to smaller inefficiency values and therefore more efficient units. However, for points closer to the origin, the denominator of (7) decreases, consequently leading to larger inefficiencies. If a DMU were to be located in the origin, then its inefficiency would measure $+\infty$.

Furthermore, also note that in the case of one input and one output, all DMUs positioned in the second and fourth quadrant are projected in a direction whose support line passes the origin. This follows directly from the choice of direction vector. Indeed, assume $z = (x, y)$ is positioned in the second quadrant. Then, $x < 0$ and $y > 0$. Consequently, $\hat{g} = (-|x|, |y|) = (x, y) = z$. From Proposition 3.1, it now follows that $z^* = z + \hat{\delta}^*\hat{g} = z + \hat{\delta}^* z = (1 + \hat{\delta}^*)z$, meaning that z^*, z, and 0 are collinear. A similar argument holds for z positioned in the fourth quadrant. This phenomenon can be observed for the points labelled 2 and 3 in Figure 1. This explains why DMU 2, which has a positive value for the single input, nevertheless has an efficiency measure larger than unity.

Thus, in the case of negative data in DEA and assuming that it is critical in practice to maintain a proportional interpretation, we recommend using the generalised proportional distance function (see Definition 3.1) and solving LP problem (6) rather than employing the standard directional distance function (Definition 2.1) or the Farrell proportional distance function (Definition 2.2) and their corresponding mathematical programs (see (2) and (3), respectively).

4. Concluding comments

The fast growing DEA literature has for a long time neglected the issues surrounding the use of negative data in managerially oriented benchmarking models. The timely work of Silva Portela *et al* (2004) suggests a variation on the directional distance function, a general distance function compatible with profit maximisation that has recently gained some popularity. This contribution has argued that a very simple modification of the traditional proportional distance function can be employed in this context instead. This generalised proportional distance function (Definition 3.1) has been shown to maintain a proportional interpretation under very mild conditions. This interpretation facilitates the use of benchmarking results in a managerial context.

Acknowledgements—We are grateful to two referees for most constructive comments. The usual disclaimer applies.

References

Ali A and Seiford L (1990). Translation invariance in data envelopment analysis. *Opns Res Lett* **9**: 403–405.

Anderson R, Lewis D and Springer T (2000). Operating efficiencies in real estate: A critical review of the literature. *J Real Estate Lit* **8**(1): 1–18.

Banker R, Charnes A and Cooper W (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Mngt Sci* **30**: 1078–1092.

Briec W (1997). A graph-type extension of Farrell technical efficiency measure. *J Prod Anal* **8**(1): 95–110.

Briec W and Kerstens K (2009). Infeasibilities and directional distance functions with application to the determinateness of the Luenberger productivity indicator. *J Optimiz Theory Appl* **141**: 55–73.

Briec W and Kerstens K (2010). Portfolio selection in multidimensional general and partial moment space. *J Econ Dynam Control* **34**: 636–656.

Chambers R (2002). Exact nonradial input, output, and productivity measurement. *Econ Theory* **20**: 751–765.

Chambers R, Chung Y and Färe R (1998). Profit, directional distance functions, and Nerlovian efficiency. *J Optimiz Theory Appl* **98**: 351–364.

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

Chavas J-P and Kim K (2007). Measurement and sources of economies of scope: A primal approach. *J Inst Theor Econ* **163**: 411–427.

Cummins D and Weiss M (2000). Analyzing firm performance in the insurance industry using frontier efficiency and productivity methods. In: Dionne G (ed). *Handbook of Insurance*. Kluwer: Boston, pp 767–829.

De Borger B, Kerstens K and Costa A (2002). Public transit performance: What does one learn from frontier studies? *Transport Rev* **22**(1): 1–38.

Emrouznejad A, Parker B and Tavares G (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio Econ Plan Sci* **42**: 151–157.

Färe R, Grosskopf S and Lovell C (1994). *Production Frontiers*. Cambridge University Press: Cambridge.

Färe R and Lovell C (1978). Measuring the technical efficiency of production. *J Econ Theory* **19**(1): 150–162.

Farrell M (1957). The measurement of productive efficiency. *J R Stat Soc Ser A-G* **120**: 253–281.

Feroz E, Kim S and Raab R (2003). Financial statement analysis: A data envelopment analysis approach. *J Opl Res Soc* **54**: 48–58.

Førsund F and Sarafoglou N (2005). The tale of two research communities: The diffusion of research on productive efficiency. *Int J Prod Econ* **98**(1): 17–40.

Harker P and Zenios S (2001). What drives the performance of financial institutions? In: Harker P and Zenios S (eds). *Performance of Financial Institutions: Efficiency, Innovation, Regulation*. Cambridge University Press: Cambridge, pp 3–31.

Luenberger D (1992). New optimality principles for economic efficiency and equilibrium. *J Optimiz Theory Appl* **75**: 221–264.

Luenberger D (1995). *Microeconomic Theory*. McGraw-Hill: Boston.

Murthi B, Choi Y and Desai P (1997). Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. *Eur J Opl Res* **98**: 408–418.

Ozcan Y (2008). *Health Care Benchmarking and Performance Evaluation: An Assessment Using Data Envelopment Analysis (DEA)*. Springer: Berlin.

Pastor J and Ruiz J (2007). Variables with negative values in DEA. In: Zhu J and Cook W (eds). *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*. Springer: Berlin, pp 63–84.

Russell R and Schworm W (2009). *Axiomatic foundations of inefficiency measurement on input, output space*. UNSW Australian School of Business Research Paper No 2009 ECON 07, http://ssrn.com/abstract=1424792.

Sharp J, Meng W and Liu W (2007). A modified slacks-based measure model for data envelopment analysis with 'natural' negative outputs and inputs. *J Opl Res Soc* **58**: 1672–1677.

Silva Portela M, Thanassoulis E and Simpson G (2004). Negative data in DEA: A directional distance approach applied to bank branches. *J Opl Res Soc* **55**: 1111–1121.

Smith P (1990). Data envelopment analysis applied to financial statements. *Omega* **18**: 131–138.

Thanassoulis E, Silva Portela M and Despić O (2008). DEA—The mathematical programming approach to efficiency analysis. In: Fried H, Lovell C and Schmidt S (eds). *The Measurement of Productive Efficiency and Productivity Change*. Oxford University Press: New York, pp 251–420.

Worthington A (2001). An empirical survey of frontier efficiency measurement techniques in education. *Educ Econ* **9**: 245–268.

Received April 2009;
accepted April 2010 after two revisions