ELSEVIER

European Journal of Operational Research 113 (1999) 206–214

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

# Estimating returns to scale using non-parametric deterministic technologies: A new method based on goodness-of-fit

Kristiaan Kerstens [a,*], Philippe Vanden Eeckaut [b]

[a] Laboratoire de Recherches Economiques et Sociales (URA 362), Université Catholique de Lille, B.P. 109, 60 Boulevard Vauban, F-59016 Lille Cédex, France

[b] CORE, Université Catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium

Received 1 August 1996; accepted 1 October 1997

**Abstract**

The purpose of this note is to define a new and more general method to obtain qualitative information about returns to scale for individual observations. The traditional methods developed for estimating returns to scale on non-parametric deterministic reference technologies (Data Envelopment Analysis (DEA) models) are reviewed. A new and more general method that is suitable for all reference technologies is provided. Its usefulness is illustrated by considering variations on an existing non-convex production model, known as the Free Disposal Hull (FDH). When different returns to scale assumptions are introduced into the FDH, then previous methods for determining returns to scale do no longer apply. 

*Keywords:* Returns to scale; FDH; DEA

## 1. Introduction

In recent years the measurement of efficiency and productivity has become a standard tool in the empirical analysis of production. [1] In particular, the possibility of fitting multiple input, multiple output correspondences using Data Envelopment Analysis (DEA) techniques (see Charnes et al., 1978) has led to an enormous amount of publications (reviewed in Seiford, 1996). In addition to evaluating the productive or technical efficiency of decision making units (DMUs), it is also possible to determine the amount of scale efficiency (SCE). Furthermore, conditional on a selected orientation of measurement one can obtain for each DMU qualitative information about scale economies. This article aims to introduce a new way to ascertain the exact source of SCEs evaluated relative to non-parametric deterministic production technologies. It is more general than existing methods since it is in fact suitable for all technologies.

* Corresponding author. Tel.: +33 3 20 1340 80; fax: +33 3 20 1340 70; e-mail: kristiaan.kerstens@flse.fupl.asso.fr

[1] Methodologies to reconstruct the boundary of the production possibility set, or one of its value duals, are extensively reviewed in Lovell (1993).

0377-2217/99/$ – see front matter 
PII: S 0 3 7 7 - 2 2 1 7 ( 9 7 ) 0 0 4 2 8 - 1

The layout of this note is as follows. Section 2 briefly reviews the traditional methods of estimating the returns to scale for individual DMUs for DEA models. In Section 3 a new and more general method that is suitable for all reference technologies is presented. In particular, when several returns to scale assumptions are introduced into a nonconvex production model, known as the Free Disposal Hull (FDH) (see Deprins et al., 1984), then the previous methods for determining returns to scale can no longer be applied. Section 4 presents the conclusion.

## 2. Estimating returns to scale in DEA models

The analysis in this section requires the definition of production technologies based on $k$ observations (DMUs) of inputs $x \in \Re^m_+$ and outputs $y \in \Re^n_+$. Technology is represented by its production possibility or transformation set $T = \{(x, y)$: $x$ can produce $y\}$, i.e., the set of all feasible input–output vectors.

Technologies differ, among others, with regard to returns to scale assumptions, i.e., the scaling of existing activities. Global scale behaviour can be defined in terms of the production possibility set $T$ (Färe et al., 1994, p. 33).

**Definition 1.** Technology exhibits Constant Returns to Scale (CRS) if $\delta T = T,\ \delta > 0$; it displays Non-Increasing Returns to Scale (NIRS) if $\delta T \subseteq T,\ 0 < \delta \leqslant 1$; and it exhibits Non-Decreasing Returns to Scale (NDRS) if $T \subseteq \delta T$, $0 < \delta \leqslant 1$.

Returns to scale can alternatively be defined in terms of individual feasible activities as follows:

**Definition 2.** Technology $T^* = \{(\delta x, \delta y)$: $\delta \in \Gamma(s)$, $\delta x$ can produce $\delta y\}$ includes the following returns to scale possibilities:
(i) $\Gamma(s) = \{\delta\colon 0 < \delta\}$ for $s =$ CRS;
(ii) $\Gamma(s) = \{\delta\colon 0 < \delta \leqslant 1\}$ for $s =$ NIRS; and
(iii) $\Gamma(s) = \{\delta\colon \delta \geqslant 1\}$ for $s =$ NDRS.

Also the following definition is useful for interpreting scale economies (Färe, 1988, p. 150).

**Definition 3.** Technology exhibits Increasing Returns to Scale (IRS) if it exhibits NDRS and not CRS. It exhibits Decreasing Returns to Scale (DRS) if it exhibits NIRS and not CRS.

Turning to the DEA models, the first reference technology, which is closest to Farrell (1957), is a constant returns to scale non-parametric deterministic production possibility set:

$$T^{\text{DEA-CRS}} = \{(x, y)\colon\ N'z \geqslant y,\ \ M'z \leqslant x,\ \ z \in \Re^k_+\},$$

where $N$ is the $k \times n$ matrix of observed outputs, $M$ is the $k \times m$ matrix of observed inputs, $z$ is a $k \times 1$ vector of intensity or activity variables, and $y$ and $x$ are $n \times 1$ and $m \times 1$ vectors of outputs, respectively, inputs. This technology, that goes back to von Neumann, imposes constant returns to scale, since there is no restriction on the intensity vector $z$, and it assumes strong input and output disposability.

Technologies assuming NIRS (NDRS) require restricting an additional constraint to be smaller (larger) or equal to unity ($I^t_k\, z \leqslant (\geqslant) 1$ where $I^t_k$ is a $k \times 1$ unity vector). [2] A final production possibility set allowing for variable returns to scale (VRS) results from the previous definition by restricting the intensity variables to sum to one ($I^t_k\, z = 1$). This technology in fact satisfies NDRS and NIRS in different regions (Färe et al., 1994).

Following Farrell (1957), efficiency is traditionally measured in a radial or equiproportional way. For the sake of convenience, attention in this paper concentrates on the radial input efficiency measure:

$$\text{DF}_i\,(x, y) = \min\{\lambda\colon\ \lambda \geqslant 0,\ \ (\lambda\, x, y) \in T\}.$$

This efficiency measure is always situated between zero and one. Efficient production on the isoquant of the technology is indicated by unity.

Technical efficiency is usually measured relative to the variable returns to scale production technology ($T^{\text{DEA-VRS}}$). Scale efficiency is evaluated relative to the constant returns to scale technology ($T^{\text{DEA-CRS}}$), since this technology provides a long

[2] The NDRS DEA model has seldom been applied (see, e.g. Seiford and Thrall, 1990, p. 16). The link between returns to scale and the convex DEA production technologies is explored further in the appendix (available upon request).

run competitive equilibrium benchmark. Efficiency measurement relative to the latter technology thus conflates scale and technical efficiencies. Therefore, it is straightforward to define scale efficiency as the ratio of two efficiency measures: one calculated on a constant returns to scale technology ($\mathrm{DF}_i(x, y \mid \mathrm{CRS})$), and one computed on a variable returns to scale technology ($\mathrm{DF}_i(x, y \mid \mathrm{VRS})$).

**Definition 4.** The input oriented scale efficiency measure ($\mathrm{SCE}_i(x, y)$) is

$$\mathrm{SCE}_i(x,y) = \mathrm{DF}_i(x,y \mid \mathrm{CRS}) \;/\; \mathrm{DF}_i(x,y \mid \mathrm{VRS}).$$

This ratio indicates the lowest possible input combination able to produce the same output in the long run as a technically efficient combination situated on the variable returns to scale technology. Since $\mathrm{DF}_i(x,y \mid \mathrm{CRS}) \leqslant \mathrm{DF}_i(x,y \mid \mathrm{VRS})$, evidently $0 < \mathrm{SCE}_i(x,y) \leqslant 1$. [3]

If $\mathrm{SCE}_i(x,y) = 1$, then the technology exhibits constant returns to scale at the observation under evaluation or at its input-oriented projection point. If $\mathrm{SCE}_i(x,y) < 1$, then it is certain that the evaluated observation is not located or projected on a piecewise linear segment where constant returns to scale prevail. In the latter case, it is possible to determine for each observation the exact nature of the returns to scale of its bounding hyperplane.

More precisely stated, three main methods have been proposed to obtain qualitative information regarding scale economies locally. [4] Each method is discussed in turn.

One way of determining the returns to scale per observation (locally) is based on summing the optimal activity vector $z^*$ on a constant returns to scale technology (Banker, 1984). As noted by Chang and Guh (1991) and Ganley and Cubbin (1992), this first method is problematic when there are alternative optimal solutions for this sum. This basically happens if there are constant returns to scale hyperplanes spanned by more than one observation. Banker and Thrall (1992) devised a method to cope with alternative optimal solutions. This method was further refined in Banker et al. (1996a, b), and Zhu and Shen (1995) who essentially try to minimise the computational burden when checking the alternative optima in the dual or multiplier LP problems.

The second method is based on inspecting the intercept of the supporting hyperplane at the reference unit for a variable returns to scale DEA technology (see Banker et al., 1984). This is simply a matter of inspecting the sign of the shadow price of the convexity constraint. Again Banker and Thrall (1992) generalised this method to deal with multiple solutions.

The third method – first proposed in Färe et al. (1983) – starts from the scale efficiency measure. It basically compares both components of the scale efficiency measure with a third efficiency measure evaluated on a technology imposing non-increasing returns to scale ($\mathrm{DF}_i(x,y \mid \mathrm{NIRS})$). [5] Since these three technologies are nested, the three input efficiency measures satisfy the following order:

$$\mathrm{DF}_i(x,y \mid \mathrm{CRS}) \leqslant \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) \leqslant \mathrm{DF}_i(x,y \mid \mathrm{VRS}).$$

On the one hand, if technical efficiency on a constant returns to scale model equals that on a non-increasing returns to scale technology ($(\mathrm{DF}_i(x,y \mid \mathrm{CRS}) = \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) < \mathrm{DF}_i(x,y \mid \mathrm{VRS}))$, then the observation is scale inefficient due to increasing returns to scale. As the lower part of the conical hull

[3] Note that this scale efficiency estimate can be made part of a more extensive efficiency decomposition based on non-parametric deterministic DEA-type models (see Färe et al., 1983, 1985, 1994).

[4] Another alternative, proposed in Førsund and Hjalmarsson (1979, 1987) but little applied, infers the average scale property of inefficient observations from a comparison of input and output efficiency measures evaluated relative to a variable returns to scale technology. The relation between scale efficiency and quantitative information regarding scale economies (i.e., scale elasticity) in multiple input multiple output DEA models is extensively discussed in Førsund (1996) and Löthgren and Tambour (1996). Recently, Banker (1996), pp. 148–151, has proposed hypothesis test statistics for testing returns to scale on DEA models.

[5] A third efficiency measure evaluated on a non-decreasing returns to scale technology could yield a similar classification. Note also that the same decomposition can be applied in a dual framework under somewhat stronger conditions: see Färe and Grosskopf (1985) and Färe et al. (1994).

(line segment 0b) is in common, the difference between the components of the scale efficiency measure can only be attributed to the fact that the observation is located or projected on an increasing returns to scale part of technology. On the other hand, if technical efficiency on the non-increasing returns to scale technology is equal to that on a variable returns to scale model $((\mathrm{DF}_i(x,y \mid \mathrm{CRS}) < \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) = \mathrm{DF}_i(x,y \mid \mathrm{VRS}))$, then the observation is scale inefficient due to decreasing returns to scale (DRS). As the upper part of the convex hull (line segment bc and beyond) is in common, the difference between the scale efficiency measure components reveals that the observation is situated or projected on that part of technology characterised by decreasing returns to scale.

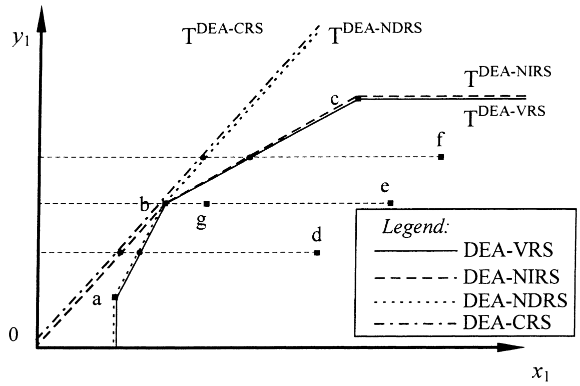


Fig. 1. Returns to scale characterisation of individual observations on DEA.

**Proposition 1.** *Using an input-oriented measurement and conditional on the optimal projection point, technology is characterised locally by:*

$$\mathrm{CRS} \Longleftrightarrow \mathrm{DF}_i(x,y \mid \mathrm{CRS}) = \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) = \mathrm{DF}_i(x,y \mid \mathrm{VRS}) \leqslant 1;$$

$$\mathrm{IRS} \Longleftrightarrow \mathrm{DF}_i(x,y \mid \mathrm{CRS}) = \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) < \mathrm{DF}_i(x,y \mid \mathrm{VRS}) \leqslant 1;$$

$$\mathrm{DRS} \Longleftrightarrow \mathrm{DF}_i(x,y \mid \mathrm{CRS}) < \mathrm{DF}_i(x,y \mid \mathrm{NIRS}) = \mathrm{DF}_i(x,y \mid \mathrm{VRS}) \leqslant 1.$$

(See Lovell (1994), p. 199. As pointed out in Färe (1997), Färe et al. (1985, 1994) in fact propose slightly different, but equivalent ways of characterising returns to scale.)

This reasoning is illustrated in Fig. 1. Observation b and the input oriented projection of observation e are clearly characterised by constant returns to scale. The observations c and f on the one hand and a and d on the other hand are subject to decreasing respectively increasing returns to scale. To spell the correct interpretation of returns to scale out in detail, we analyse observations d and f in turn. For observation d, one observes:

$$\mathrm{DF}_i(x,\ y \mid \mathrm{CRS}) = \mathrm{DF}_i(x,\ y \mid \mathrm{NIRS}) < \mathrm{DF}_i(x,\ y \mid \mathrm{VRS}).$$

While part of the conical hull is in common to both CRS and NIRS technologies, the difference with the VRS efficiency score implies that the latter technology fits best. Hence, the difference between the components of the scale efficiency measure can only be attributed to the fact that point d is located or projected on an increasing returns to scale part of the VRS technology. In case of point f it is observed that:

$$\mathrm{DF}_i(x,\ y \mid \mathrm{CRS}) < \mathrm{DF}_i(x,\ y \mid \mathrm{NIRS}) = \mathrm{DF}_i(x,\ y \mid \mathrm{VRS}).$$

Thus CRS is rejected in favour of NIRS. Following Definition 3, one can infer that observation f is subject to decreasing returns to scale.

Banker et al. (1996b) prove the equivalence of all three methods. Färe and Grosskopf (1994) stress that their method does not suffer from the problems of alternative optimal solutions encountered in the first two methods. Furthermore, Löthgren and Tambour (1996), p. 4, emphasise that the third method – in contrast to the other two methods – is not restricted to non-parametric, deterministic technologies (DEA-like models).

There are systematic informational similarities and differences of the returns to scale determination based on either input or output radial efficiency measures (see Färe et al., 1994, pp. 122, 123).[6] Input and output measures of scale efficiency (SCE) are related to each other. Since input and output technical efficiency measures calculated on a strongly disposable constant returns to scale technology are identical (equal in reciprocal terms), the ratios defining the (SCE) measures are only identical (equal in reciprocal terms) if input- and output-based efficiency measures relative to the short run technology ($T^{\text{DEA-VRS}}$) – in the denominator – are identical (equal in reciprocal terms). Otherwise, input and output measures of SCE may provide different results.

Färe et al. (1994), and more recently in Banker et al. (1996a, b), point out that the determination for each observation of the exact nature of returns to scale at its bounding hyperplane may yield conflicting information too. Both orientations predict constant, increasing and decreasing returns to scale for all observations in certain regions of the production possibility set. But for activities located in the other regions input and output projections yield contradictory information. The latter qualitative differences are caused by the fact that predictions with respect to the returns to scale of inefficient units are conditional on a move to the frontier.

## 3. A new method to determine returns to scale

An important characteristic of production that can be analysed are the determinants of returns to scale. For both observations on the frontier and in the interior of the production possibility set it is possible to determine the returns to scale of the efficient observation or the projection on the frontier. For inefficient observations, this characterisation of the returns to scale depends on the selected orientation of measurement. This limitation remains for the new method presented here.

The newly devised method is closely related to the third procedure (Färe et al., 1983) mentioned for convex production technologies in that it is based on comparing several technologies, but it systematically exploits the relationship between efficiency measures and goodness-of-fit measures used for hypothesis testing (see, e.g. Chavas and Cox, 1990; Varian, 1990).[7]

Input efficiency is computed on three different technologies imposing respectively constant, non-increasing and non-decreasing returns to scale while maintaining a common set of technological assumptions. These three technologies are not nested and embody respectively the null hypothesis of constant, decreasing, and increasing returns to scale. If the efficiency measure evaluated relative to any of these three technologies is unity, then the embodied null hypothesis cannot be rejected. If this efficiency measure is smaller than unity, then it indicates the relative strength of the evidence against the null hypothesis. Therefore, selecting the maximal efficiency measure among the three technologies amounts to minimising the strength of rejection of the embodied null hypotheses. Positively stated, it identifies the null hypothesis that is most compatible with the observations.

Formally, it is possible to infer for any single observation whether it satisfies constant (CRS), increasing (IRS), or decreasing returns (DRS) to scale by simply identifying the technology yielding the maximal input efficiency score.

**Proposition 2.** *Using an input-oriented measurement and conditional on the optimal projection point, technology is characterised locally by:*

$$\text{CRS} \iff \text{DF}_i(x,y \mid \text{CRS}) = \max\{\text{DF}_i(x,y \mid \text{CRS}), \text{DF}_i(x,y \mid \text{NIRS}), \text{DF}_i(x,y \mid \text{NDRS})\};$$

$$\text{IRS} \iff \text{DF}_i(x,y \mid \text{NDRS}) = \max\{\text{DF}_i(x,y \mid \text{CRS}), \text{DF}_i(x,y \mid \text{NIRS}), \text{DF}_i(x,y \mid \text{NDRS})\}; \quad \text{or}$$

[6] Similar remarks hold true for the less often utilised graph efficiency measures (that simultaneously aim to reduce input usage and expand the level of outputs).

[7] Färe and Grosskopf (1995) explore the relation between goodness-of-fit tests and efficiency measures in detail.

$$\text{DRS} \Longleftrightarrow \text{DF}_i(x,y \mid \text{NIRS}) = \max\{\text{DF}_i(x,y \mid \text{NDRS})\}.\text{DF}_i(x,y \mid \text{CRS}), \text{DF}_i(x,y \mid \text{NIRS}).$$

(Note that all three input efficiency measures coincide for observations subject to constant returns to scale. Remark also that the same formulas apply for any efficiency measure smaller than unity. Obviously, for radial output efficiency measures defined to be larger than unity the max operator should be replaced by a min operator.)

The maximal input efficiency measure simply reflects the best fit of a specific technology for the given observation and therefore serves to indicate the most appropriate returns to scale assumption. This procedure can be applied to any non-parametric, deterministic frontier model, including the convex DEA models. In fact, it is applicable to any specification of technology (Löthgren and Tambour, 1996). This new procedure is therefore more general.

To explicate the relation between our new method and the procedure of Proposition 1 we first note the following preliminaries. The DEA constant returns to scale technologies can be seen as the union of its non-increasing and non-decreasing returns to scale technologies ($T^{\text{DEA-CRS}} = T^{\text{DEA-NIRS}} \cup T^{\text{DEA-NDRS}}$). Likewise, the DEA variable returns to scale model equals the intersection of its non-increasing and non-decreasing returns to scale counterparts ($T^{\text{DEA-VRS}} = T^{\text{DEA-NIRS}} \cap T^{\text{DEA-NDRS}}$).

To fix our ideas we relate the second lines of both formulas. Under the initial method

$$\text{DF}_i(x,\ y \mid \text{CRS}) = \text{DF}_i(x,\ y \mid \text{NIRS}) < \text{DF}_i(x,\ y \mid \text{VRS}) \leqslant 1$$

allows to conclude that the observation considered is subject to IRS. Since input efficiency measured relative to $T^{\text{DEA-CRS}}(= T^{\text{DEA-NIRS}} \cup T^{\text{DEA-NDRS}})$ equals input efficiency measured relative to $T^{\text{DEA-NIRS}}$, it is clear that $T^{\text{DEA-NDRS}}$ does not add anything to the construction of $T^{\text{DEA-CRS}}$ locally. In Fig. 1 we are at that part of the boundary where $T^{\text{DEA-CRS}}$ and $T^{\text{DEA-NIRS}}$ coincide. Furthermore, since input efficiency measured relative to $T^{\text{DEA-NIRS}}$ is smaller than input efficiency measured relative to $T^{\text{DEA-VRS}}(= T^{\text{DEA-NIRS}} \cap T^{\text{DEA-NDRS}})$, we infer that $T^{\text{DEA-NDRS}}$ determines the intersection locally. In terms of Fig. 1, we are situated at that part of the boundary where $T^{\text{DEA-VRS}}$ and $T^{\text{DEA-NDRS}}$ coincide. Hence we conclude that input efficiency measured relative to $T^{\text{DEA-NDRS}}$ (which in this case equals $\text{DF}_i(x,\ y \mid \text{VRS})$) yields the maximum efficiency score and thus the observation is situated at the IRS part of technology.

Thus, in the Färe et al. (1983) method the comparison with the $T^{\text{DEA-NDRS}}$ technology is implicitly (by means of the $T^{\text{DEA-VRS}}$ model), while our method makes this comparison explicitly.

The use of the new method is illustrated by considering a series of non-convex technologies some of which have been proposed in Bogetoft (1996). He convincingly argues that it may be useful to deconvexify the production possibility set to model the advantages of specialisation in the production of output or in the use of inputs. He therefore proposed, among others, several generalisations of an existing non-convex frontier known as the FDH. We first specify the traditional FDH technology, as defined in Deprins et al. (1984) and applied in Bauer and Hancock (1993), Fried et al. (1995), Kerstens (1996), Lovell (1995), among others. Then the alternative non-convex production models imposing additional returns to scale assumptions are defined. [8]

First, the FDH technology can be represented by its production possibility set

$$T^{\text{FDH}} = \Big\{(x,y):\ N'z \geqslant y,\ M'z \leqslant x,\ I_k'z = 1,\ z_i \in \{0,1\}\Big\}.$$

FDH only imposes strong disposability assumptions. It is worth stressing that FDH does not impose any specific returns to scale hypothesis, defined previously. [9] Radial technical efficiency

[8] Properties of some of these technologies are discussed in Bogetoft (1996), p. 464.

[9] In particular, FDH does not impose VRS as defined above, despite the superficial similarity with the VRS DEA model (i.e., the common constraint $I_k{}^{t}z = 1$).

measurement requires solving a mixed integer linear programming problem for each DMU. [10] As shown in Tulkens (1993), enumeration algorithms based upon vector dominance reasoning can be used.

It is possible to add to the previous production model a particular returns to scale assumption: i.e., constant, non-increasing, and non-decreasing returns to scale. Modelling constant returns to scale results in the following technology:

$$T^{\text{FDH-CRS}} = \Big\{(x,y):\ N'w \geqslant y,\ M'w \leqslant x,\ I_k'z = 1,\ z_i \in \{0,1\},\ w_i = \delta z_i,\ \delta \geqslant 0\Big\}.$$

There is now one activity vector $z$ operating subject to a non-convexity constraint and one rescaled activity vector $w$ allowing for any scaling of the observations spanning the frontier. The scaling parameter ($\delta$) is free.

Non-increasing returns to scale is imposed by adding an additional restriction on the scaling parameter $\delta$ to the previous non-convex constant returns to scale technology:

$$T^{\text{FDH-NIRS}} = \Big\{(x,y):\ N'w \geqslant y,\ M'w \leqslant x,\ I_k'z = 1,\ z_i \in \{0,1\},\ w_i = \delta z_i,\ 0 \leqslant \delta \leqslant 1\Big\},$$

where the scaling parameter $\delta$ is constrained to be smaller than or equal to unity. Non-decreasing returns to scale results from restraining the scaling parameter $\delta$ to be larger than or equal to unity:

$$T^{\text{FDH-NDRS}} = \Big\{(x,y):\ N'w \geqslant y,\ M'w \leqslant x,\ I_k'z = 1,\ z_i \in \{0,1\},\ w_i = \delta z_i,\ \delta \geqslant 1\Big\}.$$

Non-increasing and non-decreasing returns to scale allow for a lower respectively an upper proportionality of observed activities by means of the scaling parameter ($\delta$).

[10] Since the radial efficiency measure evaluates performance relative to the isoquant, and not to the much smaller efficient subset, it has been argued that non-radial efficiency measures provide a better alternative (see, e.g. De Borger and Kerstens, 1996).

Fig. 2 shows the graph of these three non-convex technologies together with the traditional FDH in a single input single output space. Notice that $T^{\text{FDH-NIRS}}$ and $T^{\text{FDH-NDRS}}$ are related to technologies proposed in Petersen (1990), except that this author maintains the convexity assumption in input space and in output space. In the single input single output case, however, $T^{\text{FDH-NIRS}}$ and $T^{\text{FDH-NDRS}}$ are identical to the Petersen (1990) technologies.

Crucial is that the traditional methods to determine returns to scale no longer apply. The first two methods for the obvious reason that duality relations break down for mixed integer programming problems. The third procedure only works in case the variable returns to scale model in fact incorporates all information contained in the non-increasing and non-decreasing returns to scale technology. While the FDH constant returns to scale technology is still the union of non-increasing and non-decreasing returns to scale FDH technologies ($T^{\text{FDH-CRS}} = T^{\text{FDH-NIRS}} \cup T^{\text{FDH-NDRS}}$), it is not true that the traditional FDH model equals the intersection of its non-increasing and non-decreasing returns to scale technologies ($T^{\text{FDH}} \subseteq T^{\text{FDH-NIRS}} \cap T^{\text{FDH-}NDRS}$). This can be illustrated by contrasting observations d and f. While in the case of d the projection point on the constant and non-increasing returns to scale technologies coincide, for observation f all three projection points are different and no conclusion is possible.

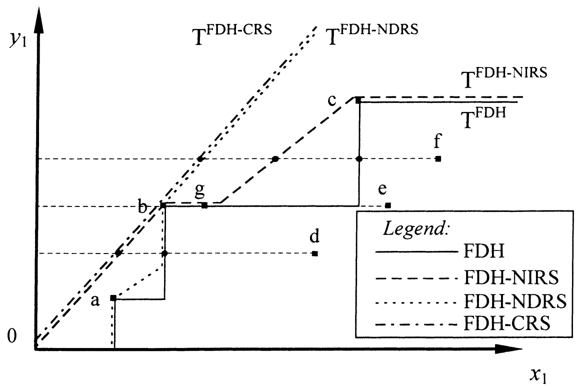


Fig. 2. Returns to scale characterisation of individual observations on non-convex non-parametric deterministic frontiers.

Illustrating our new approach it is clear that observation b and the projection of e are compatible with the constant returns to scale hypothesis, while observations c and f on the one hand and observations a and d on the other hand are on the decreasing respectively increasing parts of the technology. Again, changing the measurement orientation may lead to diverging conclusions for DMUs. Unit d, for instance, enjoys increasing and decreasing returns to scale under input respectively output measurement, while observation g experiences constant and decreasing scale economies under the same orientations.

While the new method implies a higher computational cost compared to existing methods, it is evident that it is attractive in both its generality and clarity.

These non-convex reference technologies again yield a series of models allowing to determine scale efficiency ($\mathrm{SCE}_i(x, y)$). The relation between convex and non-convex scale efficiencies requires some clarification. Observe that the non-convex technologies are nested in the convex technologies constituting the SCE measure:

$$T^{\text{FDH}} \subseteq T^{\text{DEA-VRS}};$$

$$T^{\text{FDH-CRS}} \subseteq T^{\text{DEA-CRS}}.$$

While it follows that the underlying efficiency measures can be ordered, it is impossible to order the ratios between these efficiency measures. Consequently, there is no a priori ordering between both SCE measures. [11] Furthermore, any difference between efficiency measures evaluated relative to convex and non-convex technologies with otherwise identical returns to scale assumptions can be completely attributed to the convexity assumption. This provides a perfect base to compare the impact of the convexity assumption.

We close with a suggestion to reduce the effect of measurement orientation. One could check for inefficient observations $(x^{\circ},y^{\circ})$ the returns to scale of all elements in their set of dominating observations $(B(x^{\circ},y^{\circ}) = \{(x_k,y_k)\colon\ x_k \leqslant x^{\circ},\ y_k \geqslant y^{\circ}\})$. If all dominating observations in the non-convex case or all faces in the convex case are subject to the same returns to scale, then the stability of the conclusions increases.

[11] All the efficiency computations on these three non-convex technologies also require solving mixed integer non-linear programming problems for each observation being evaluated. As pointed out in Bogetoft (1996) and as shown in an appendix (available upon request) it is possible to make use of enumeration algorithms based upon vector dominance reasoning. Kerstens and Vanden Eeckaut (1997) present an empirical application based on these non-convex technologies.

## 4. Conclusion

Starting from a review of existing ways of estimating SCE and in particular returns to scale using convex (DEA) non-parametric production models, a new and more general way of determining returns to scale for both efficient and inefficient observations has been proposed. This new method exploits the relation between efficiency measures and goodness-of-fit tests. It is indispensable, since traditional methods cannot be applied to a series of non-convex production models with different returns to scale assumptions. Inevitably, the returns to scale determination remains conditional on the choice of a measurement orientation.

## Acknowledgements

This paper benefited from comments by R. Färe, H. Tulkens and three anonymous referees. The usual disclaimer applies.

## References

Banker, R., 1984. Estimating most productive scale size using data envelopment analysis. European Journal of Operational Research 17 (1), 35–44.

Banker, R., 1996. Hypothesis test using data envelopment analysis. Journal of Productivity Analysis 7 (2/3), 139–159.

Banker, R., Thrall, R., 1992. Estimation of returns to scale using data envelopment analysis. European Journal of Operational Research 62 (1), 74–84.

Banker, R., Bardhan, I., Cooper, W., 1996a. A note on returns to scale in DEA. European Journal of Operational Research 88 (3), 583–585.

Banker, R., Chang, H., Cooper, W., 1996b. Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis. European Journal of Operational Research 89 (3), 473–481.

Banker, R., Charnes, A., Cooper, W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092.

Bauer, P., Hancock, D., 1993. The efficiency of the federal reserve in providing check processing services. Journal of Banking and Finance 17 (2/3), 287–311.

Bogetoft, P., 1996. DEA on relaxed convexity assumptions. Management Science 42 (3), 457–465.

Chang, K.-P., Guh, Y.-Y., 1991. Linear production functions and the data envelopment analysis. European Journal of Operational Research 52 (2), 215–223.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429–444.

Chavas, J.-P., Cox, T., 1990. A non-parametric analysis of productivity: The Case of US and Japanese manufacturing. American Economic Review 80 (3), 450–464.

De Borger, B., Kerstens, K., 1996. Radial and nonradial measures of technical efficiency: an empirical illustration for Belgian local governments using an FDH reference technology. Journal of Productivity Analysis 7 (1), 41–62.

Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labor-efficiency in post offices, in: Marchand, M., Pestieau, P., Tulkens, H. (Eds.), The Performance of Public Enterprises: Concepts and Measurement. Elsevier, Amsterdam, pp. 243–267.

Färe, R., 1988. Fundamentals of Production Theory. Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin.

Färe, R., 1997. Determining Returns to Scale in DEA, Discussion Paper 97-10, Department of Economics, Southern Illinois University, Carbondale, IL.

Färe, R., Grosskopf, S., 1985. A nonparametric cost approach to scale efficiency. Scandinavian Journal of Economics 87 (4), 594–604.

Färe, R., Grosskopf, S., 1994. Estimation of returns to scale using data envelopment analysis: A Comment. European Journal of Operational Research 79 (3), 379–382.

Färe, R., Grosskopf, S., 1995. Nonparametric tests of regularity, Farrell efficiency and goodness-of-fit. Journal of Econometrics 69 (2), 415–425.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1983. The structure of technical efficiency. Scandinavian Journal of Economics 85 (2), 181–190.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1985. The Measurement of Efficiency of Production. Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1994. Production Frontiers. Cambridge University Press, Cambridge.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A, 120 (3), 253–281.

Førsund, F., Hjalmarsson, L., 1979. Generalised Farrell measures of efficiency: An application to milk processing in Swedish dairy plants. Economic Journal 89 (354), 294–315.

Førsund, F., Hjalmarsson, L., 1987. Analyses of Industrial Structure: A Putty–Clay Approach, Almqvist & Wiksell, Stockholm.

Førsund, F., 1996. On the calculation of scale elasticity in DEA models. Journal of Productivity Analysis 7 (2/3), 283–302.

Fried, H.O., Lovell, C.A.K., Vanden Eeckaut, P., 1995. Service productivity in US credit unions, in: Harker, P.T. (Ed.), The Service Productivity and Quality Challenge. Kluwer Academic Publishers, Boston, pp. 365–390.

Ganley, J., Cubbin, J., 1992. Public Sector Efficiency Measurement: Applications of Data Envelopment Analysis. North-Holland, Amsterdam.

Kerstens, K., 1996. Technical efficiency measurement and explanation of French urban transit companies. Transportation Research 30A (6), 431–452.

Kerstens, K., Vanden Eeckaut, P., 1997. Distinguishing Technical and Scale Efficiency on Non-Convex and Convex Technologies: An Illustration Using UK Rates Departments, Université Catholique de Lille, Lille (LABORES: mimeo).

Löthgren, M., Tambour, M., 1996. Alternative approaches to estimate returns to scale in DEA-models, Working paper 90, Economic Research Institute, Stockholm School of Economics, Stockholm.

Lovell, C.A.K., 1993. Production frontiers and productive efficiency, in: Fried, H., Lovell, C.A.K., Schmidt S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, Oxford, pp. 3–67.

Lovell, C.A.K., 1994. Linear programming approaches to the measurement and analysis of productive efficiency. TOP 2 (2), 175–224.

Lovell, C.A.K., 1995. Measuring the macroeconomic performance of the Taiwanese economy. International Journal of Production Economics 39 (1), 165–178.

Petersen, N., 1990. Data envelopment analysis on a relaxed set of assumptions. Management Science 36 (3), 305–314.

Seiford, L., Thrall, R., 1990. Recent developments in DEA: the mathematical programming approach to frontier analysis. Journal of Econometrics 46 (1/2), 7–38.

Seiford, L., 1996. Data envelopment analysis: the evolution of the state of the art (1978–1995). Journal of Productivity Analysis 7 (2/3), 99–137.

Tulkens, H., 1993. On FDH efficiency analysis: some methodological issues and applications to retail banking, courts, and urban transit. Journal of Productivity Analysis 4 (1/2), 183–210.

Varian, H., 1990. Goodness-of-fit in optimizing models. Journal of Econometrics 46 (1/2), 125–140.

Zhu, J., Shen, Z.-H., 1995. A discussion of testing DMU's returns to scale. European Journal of Operational Research 81 (3), 590–596.