Journal of Productivity Analysis, 22, 227–257, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

Efficiency and Global Scale Characteristics on the "No Free Lunch" Assumption Only

VICTOR V. PODINOVSKI v.podinovski@warwick.ac.uk
Warwick Business School, University of Warwick, Coventry CV4 7AL, UK

Abstract

In a production technology, the type of returns to scale (RTS) associated with an efficient decision making unit (DMU) is indicative of the direction of marginal rescaling that the DMU should undertake in order to improve its productivity. In this paper a concept of global returns to scale (GRS) is developed as an indicator of the direction in which the most productive scale size (MPSS) of an efficient DMU is achieved. The GRS classes are useful in assisting strategic decisions like those involving mergers of units or splitting into smaller firms. The two characterisations, RTS and GRS, are the same in a convex technology but generally different in a non-convex one. It is shown that, in a non-convex technology, the well-known method of testing RTS proposed by Färe et al. is in fact testing for GRS and not RTS. Further, while there are three types of RTS: constant, decreasing and increasing (CRS, DRS and IRS, respectively), the classification according to GRS includes the fourth type of sub-constant GRS, which describes a DMU able to achieve its MPSS by both reducing and increasing the scale of operations. The notion of GRS is applicable to a wide range of technologies, including the free disposal hull (FDH) and all polyhedral technologies used in data envelopment analysis (DEA).

JEL Classification: C61, C67

Keywords: efficiency, returns to scale, most productive scale size, data envelopment analysis, free disposal hull

1. Introduction

The concept of returns to scale has been studied extensively in production economics (see, e.g., Färe et al., 1983, 1985). On the practical side, the returns to scale (RTS) classification is often used in data envelopment analysis, most often in relation to the BCC model of Banker et al. (1984). For an overview of the results in this area see, e.g., Seiford and Zhu (1999), Cooper et al. (2000), Thanassoulis (2001) and Banker et al. (2004).

The classes of RTS are indicative of the type of *marginally small* resizing that an efficient unit should undertake in order to see an immediate improvement of its average productivity. For example, if a unit exhibits increasing returns to scale (IRS), it should benefit by marginally increasing the scale of its operations, provided it remains located on the efficient frontier.

As shown by Banker (1984), in the polyhedral production technology of the BCC model, the type of RTS also serves as *a global indicator* of the direction towards the most productive scale size (MPSS). More precisely, if a decision making unit (DMU) exhibits constant returns to scale (CRS), it operates at MPSS. If it exhibits IRS, it does not operate at MPSS but would achieve it by scaling its operations up. If the DMU exhibits decreasing returns to scale (DRS), it would achieve its MPSS by scaling its operations down.

This dual role of the RTS classification as a local improvement indicator and direction to MPSS is preserved in any convex technology,[1] which is formally established in Theorem 7 below. However, in a general non-convex technology the RTS classes no longer play a role of global indicators.[2] This observation is illustrated by a number of examples throughout this paper.

The method developed by Färe et al. defines the type of RTS exhibited by a particular DMU by comparing its radial efficiency in the production technology T with its efficiency in the CRS, or cone, reference technology. If the two are equal, the DMU is said to exhibit CRS. Otherwise there is scale inefficiency, and the non-increasing returns-to-scale (NIRS) reference technology is used to investigate its nature. If the radial efficiency of the DMU in the NIRS technology is not equal to that in the CRS technology, the scale inefficiency is explained by the fact that the DMU is exhibiting DRS, otherwise the reason of scale inefficiency must be IRS. A modification of this method was developed by Kerstens and Vanden Eeckaut (1999).

Neither Färe et al. nor Kerstens and Vanden Eeckaut assume that the production technology T is convex. In this paper we show that this assumption is, however, essential for obtaining the correct classification of RTS by the above methods.

Further, we show that, if the technology T is not convex, the types of RTS produced by these methods, although generally incorrect, are still "almost" correct as the global characteristics indicating the direction to MPSS. In other words, the above methods test for the types of *global* resizing guiding an efficient DMU to its MPSS. This leads to the concept of *global returns to scale* (GRS), which was first investigated in Podinovski (2002, 2004b).

The fact that the methods of Färe et al. and Kerstens and Vanden Eeckaut do not produce the correct classification of RTS in non-convex technologies should not be surprising. The obvious fundamental problem here is that the type of RTS exhibited by a unit is fully defined by its *marginally small* neighbourhood. At the same time, the two methods relate the type of RTS to the efficiency of DMU_0 in the reference technologies, which are generally defined by some "global players", like the MPSS units in the CRS reference technology. We can therefore conclude that the reference technology methods are suitable for testing local RTS in a convex technology only because, in such a technology, the local and global indicators of improvement coincide. In a general non-convex technology, these methods only test for the global indicators.

In this paper, we develop the notion of GRS under an *extremely weak* set of assumptions about the production technology. Namely, we only assume a variant of the "no free lunch" assumption, which states that positive outputs cannot be produced from a zero input vector.

The notion of MPSS is the starting point in this development. We show that for any efficient DMU there are only four options, regarding the types of resizing leading to its MPSS. These four options become the basis of the definition of GRS as an indicator of the direction in which the MPSS can be achieved. In a convex technology, three of these options coincide with the traditional CRS, DRS and IRS classes of local RTS, but in a non-convex technology the classification is generally different.[3] More, in a non-convex technology, the fourth type of GRS, which we call *sub-constant*, can also be identified. This corresponds to the case in which an efficient DMU can achieve its MPSS in two ways: by scaling its operations down or, alternatively, up.

From the managerial perspective, the four types of GRS should be useful in aiding strategic decisions like those concerning merging or splitting of production units, where the notion of local RTS could be misleading. It is also worth emphasizing that the GRS classification can be applied to virtually any technology, including the free disposal hull (FDH) (Deprins et al., 1984), where the local RTS classification is, strictly speaking, undefined.

In order to test for GRS, we modify the method of Färe et al. by replacing the CRS by the non-decreasing returns-to-scale (NDRS) reference technology in the troika of technologies used in the method. We are then able to prove that the four different configurations, which the efficiencies in the troika can form, correspond to the four types of GRS.

We proceed as follows. The basic assumptions and definitions of reference technologies are given in Section 2. The concepts of efficiency and MPSS are then defined in Sections 3 and 4. In Sections 5–7, starting with illustrative examples, the concept of GRS is developed. In Section 8, the relation between the GRS and RTS is discussed. The relevance of the GRS classes to the neoclassical notion of multiproduct scale economies is also outlined. A method of testing GRS is then introduced in Section 9. The concept of GRS for inefficient units is discussed in Section 10. In Section 11, the GRS and RTS classifications are compared in the case of a convex technology.

2. Definitions

Let T be the PPS of a production technology with m inputs and s outputs. Any member of T is referred to as a DMU and represented by a pair (X, Y), where X is the m-dimensional vector of inputs and Y the s-dimensional vector of outputs. The set T is assumed to be a non-empty subset of the non-negative orthant R_{+}^{m+s}.

For any X, define $P(X)=\{Y|(X,Y)\in T\}$, the set of all output vectors Y obtainable from X in technology T. Similarly, for any Y define $L(Y)=\{X|(X,Y)\in T\}$, the set of all input vectors X, from which Y can be obtained.

Throughout this paper the set T is assumed to satisfy the following two assumptions, where $\|\cdot\|$ is the Euclidean norm:

(A1) For any X, either $P(X) = \emptyset$ or $\sup\{\|Y\| \mid Y \in P(X)\} < +\infty$.

(A2) For any $Y \neq 0$, either $L(Y) = \emptyset$ or $\inf\{\|X\| \mid X \in L(Y)\} > 0$.

The assumption (A1) ensures that there is a finite upper bound on the level of outputs, which can be produced from any given vector of inputs. Similarly, the assumption (A2) ensures that there is a non-zero lower bound on the level of inputs necessary to produce any non-zero vector of outputs. This obviously implies that no positive outputs can be produced from zero inputs. The latter is often referred to as the "no free lunch" assumption. Further discussion on this theme is included at the end of this section.

Throughout this paper, the following three reference technologies generated by T are frequently used:

$$C = \{(X, Y) \mid (X, Y) = \alpha(X', Y');\ (X', Y') \in T, \alpha \geq 0\},$$
$$H = \{(X, Y) \mid (X, Y) = \alpha(X', Y');\ (X', Y') \in T;\ 0 \leq \alpha \leq 1\},$$
$$G = \{(X, Y) \mid (X, Y) = \alpha(X', Y');\ (X', Y') \in T;\ \alpha \geq 1\}.$$

Although C, H and G are generally neither convex nor closed, we preserve their conventional names used in the convex case. Thus C is referred to as the CRS, H as the NIRS, and G as the NDRS technologies.

Note that technologies C, H and G may not satisfy the assumptions (A1) and (A2), even though this is assumed about T. In Figure 1, the convex PPS T is defined by the correspondence $0 \leq Y \leq \sqrt{X}$, and the reference technologies C and G coincide with the set $R^2_+ \backslash OY_+$, where OY_+ is the strictly positive part of the OY axis. Thus C and G do not satisfy (A1) and (A2).

In Figure 2, the non-convex T is defined as $0 \leq Y \leq X^2$. Clearly, the technologies C and H coincide with $R^2_+ \backslash OY_+$ and do not satisfy (A1) and (A2).

Our final remark is of purely academic interest. Let $\bar{C}$ be the closure of C. Then $\bar{C}$ is the minimal closed (possibly non-convex) cone technology generated by T. Consider the following statement ("no free lunch" in $\bar{C}$):

$$[(X, Y) \in \bar{C} \text{ and } X = 0] \quad \text{implies } Y = 0. \tag{1}$$

Lemma 1. *If statement* (1) *is true, then* T *satisfies* (A1) *and* (A2).

It is interesting to note that the converse to Lemma 1 is not true. (A counter-example is the technology T in Figure 1.) Therefore, the assumptions (A1) and (A2) together are weaker than (1). This completely justifies the title of our paper.

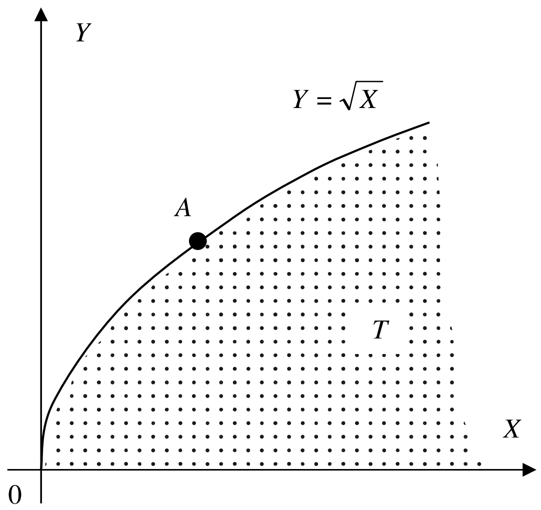

Figure 1. Infinite productivity at the origin.

3. Efficiency

Most results in this section may appear to be a repetition of well-known facts. Our purpose is, however, to show that these definitions, models and approaches remain valid under the extremely weak assumptions (A1) and (A2) only, which is a less obvious statement. This is in line with the overall approach adopted in this paper, where only these two conditions are assumed unless otherwise stated.

The concept of efficiency in DEA is so basic that even the weak assumptions (A1) and (A2) about the PPS T are not needed for its definition.

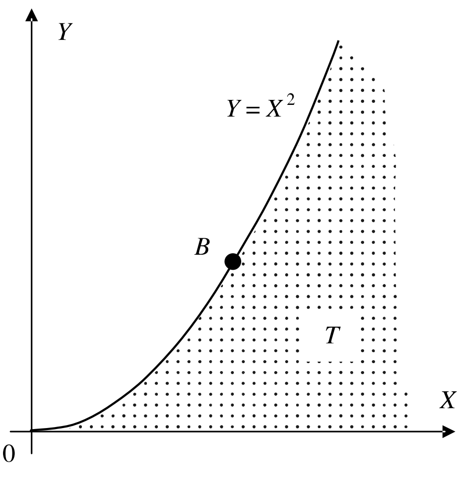

Figure 2. Infinite productivity at infinity.

Definition 1. DMU (X, Y) is called efficient if it is not dominated by any other DMU from T, that is, if $(X', Y') \in T$, then $X' \leq X$ and $Y' \geq Y$ implies $X' = X$ and $Y' = Y$. Otherwise, DMU (X, Y) is called inefficient.

Let $\text{DMU}_0 = (X_0, Y_0)$ be any DMU from T such that $X_0 \neq 0$ and $Y_0 \neq 0$. (This convention about DMU_0 is maintained throughout this paper without mention.) The input radial efficiency of DMU_0 is defined as $E_T^{\text{i}}(X_0, Y_0) = \theta^{\text{i}}$, where

$$\theta^{\text{i}} = \inf \theta \quad \text{subject to } (\theta X_0, Y_0) \in T. \tag{2}$$

Similarly, the output radial efficiency of DMU_0 is defined as $E_T^{\text{o}}(X_0, Y_0) = 1/\theta^{\text{o}}$, where

$$\theta^{\text{o}} = \sup \theta \quad \text{subject to } (X_0, \theta Y_0) \in T. \tag{3}$$

Both programs (2) and (3) are obviously feasible. Due to assumptions (A1) and (A2), we have $0 < \theta^{\text{i}} \leq 1$ and $1 \leq \theta^{\text{o}} < +\infty$. Consequently, in both cases, $0 < E_T^{\text{o}}(X_0, Y_0) \leq 1$.

Substituting T by C, H and G in models (2) and (3) we obtain, respectively, the input radial efficiency measures $E_C^{\text{i}}(X_0, Y_0)$, $E_H^{\text{i}}(X_0, Y_0)$ and $E_G^{\text{i}}(X_0, Y_0)$, and output radial efficiency measures $E_C^{\text{o}}(X_0, Y_0)$, $E_H^{\text{o}}(X_0, Y_0)$ and $E_G^{\text{o}}(X_0, Y_0)$.

As noted, technologies C, H and G may not satisfy (A1) and (A2). Therefore, with T substituted by C, H or G in model (2), the infimum θ^{i} and, consequently, the input radial efficiency of DMU_0 in the respective technology may be equal to zero. Similarly, in model (3) θ^{o} may be equal to $+\infty$, in which case the corresponding output efficiency of DMU_0 is also equal to zero.

In the case of any of the technologies T, C, H or G, the infimum θ^{i} in (2) and supremum θ^{o} in (3) are not generally attained. If, however, $\theta^{\text{i}} = 1$, it is obviously attained because $\text{DMU}_0 \in T$. Similarly, if $\theta^{\text{o}} = 1$, it is also attained.

THEOREM 1.

(i) *for any* $\text{DMU}_0 \in T$, $E_C^{\text{i}}(X_0, Y_0) = E_C^{\text{o}}(X_0, Y_0)$,

(ii) *if* DMU_0 *is efficient in* T, *then* $E_H^{\text{i}}(X_0, Y_0) = E_H^{\text{o}}(X_0, Y_0)$ *and* $E_G^{\text{i}}(X_0, Y_0) = E_G^{\text{o}}(X_0, Y_0)$.

As an unconventional illustration to Theorem 1, consider the efficient DMU A in Figure 1. Clearly, the input and output radial efficiencies of this DMU in technology $H = T$ are equal to 1. At the same time, both radial efficiencies in technologies C and G are equal to zero.

Theorem 1 allows us to use the same generic notation $E_C(X_0, Y_0)$ for both the input and output radial efficiencies of DMU_0 in technology C.

To test whether DMU_0 is efficient in the sense of Definition 1 and, if not, identify its efficient target, we may perform the second optimisation stage in which the sum of component slacks is maximised (Ali and Seiford, 1993). Although this technique was originally developed for the CCR (Charnes et al., 1978) and BCC models, it is easy to see that it remains robust under the weak assumptions (A1) and (A2). Further details of this can be found in Podinovski (2002).

4. The Most Productive Scale Size

Following Banker (1984), the most productive scale size for DMU_0 can be defined in a way, which is invariant with respect to the orientation of the model.

Definition 2. DMU_0 operates at MPSS if for all DMUs $(\delta X_0, \gamma Y_0) \in T$, where δ, $\gamma > 0$, the ratio $\delta/\gamma \geq 1$.

Clearly, DMU_0 operates at MPSS if and only if the infimum in the following program is equal to 1 (in which case it is attained at $\delta = \gamma = 1$):

$$m^* = \inf \quad \delta/\gamma \tag{4}$$
$$\text{subject to} \quad (\delta X_0, \gamma Y_0) \in T, \quad \delta > 0, \quad \gamma > 0.$$

Banker (1984) uses an equivalent formulation to (4), in which the reciprocal ratio γ/δ is maximised. We prefer (4) because its optimum is always finite, which ensures notational uniformity and reduces the amount of technical detail in our development.

Noting that the infimum m^* in (4) is not changed if T is replaced by C and γ set equal to 1, we obtain

$$m^* = E_C(X_0, Y_0).$$

If DMU_0 operates at MPSS, its input and output radial efficiency in technology C, and therefore in T, is equal to 1. DMU_0 may, however, be inefficient in these technologies in the sense of Definition 1 because of possible non-proportional improvements (residual slacks) in its inputs and outputs. Such DMUs exhibit mix inefficiency and are specifically excluded from the definition of MPSS by Cooper et al. (2000). We are not, however, excluding such DMUs from the definition of MPSS (see Section 10 for a further discussion of this issue).

5. Global Returns to Scale: Motivational Examples

In this section, we outline the concept of GRS using two examples: one of an FDH technology and the other involving its smooth analogue.

Consider the single input and single output FDH technology T in Figure 3. There are six efficient DMUs: A, B, D, E, F and K.

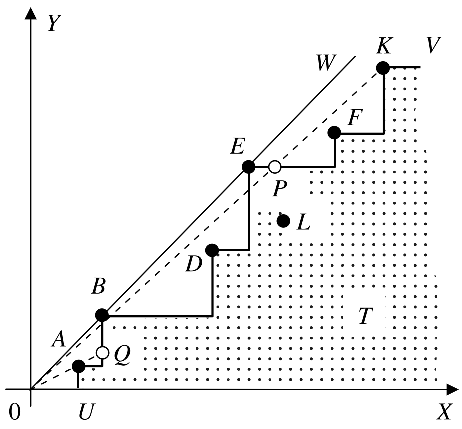

Figure 3. An FDH technology.

Define the average productivity of any DMU $(X, Y) \in T$ as the ratio Y/X. We now constrain our attention to the horizontal parts of the boundary of T in Figure 3, that is to those DMUs (X, Y) in which Y is the maximum level of output obtainable from X. The average productivity of such DMUs is a function of one variable X only. This function will be denoted as $\varphi(X)$ and is shown in Figure 4, where φ^* is the maximum of $\varphi(X)$.

The six efficient DMUs in T can now be referred to the following four different categories, which are logically exhaustive.

(1) The average productivity of DMUs b and E is equal to its maximum φ^*.This is equivalent to the fact that B and E operate at MPSS. We shall say that these two DMUs exhibit *global constant RTS* (G-CRS).

(2) DMUs F and K are larger than any DMU, at which φ^* is attained. This means that DMUs F and K will have to reduce the scale of operations to achieve MPSS. We shall say that these two DMUs exhibit *global decreasing RTS* (G-DRS).

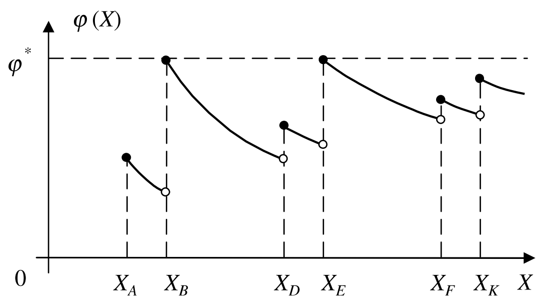

Figure 4. Average productivity $\phi(X)$ as a function of X.

(3) DMU A is smaller than any DMU, at which φ^* is attained. Consequently, A has to increase the scale of operations to achieve MPSS and therefore exhibits *global increasing RTS* (G-IRS).

(4) The average productivity of DMU D is less than the maximum φ^*, and D is between the units B and E at which φ^* is attained. This means that DMU_0 does not operate at MPSS but can choose whether to increase or reduce the scale of its operations in order to achieve MPSS. We shall say that DMU D exhibits *global sub-constant RTS* (G-SCRS).

These four types of GRS are not, of course, specific to the FDH technology. In Figure 5, a smooth variant of the previous technology is presented. According to the GRS classification, the DMUs B and E operate at MPSS and exhibit G-CRS. Any DMU on the boundary above E, which includes F and K, exhibits G-DRS. Any DMU on the boundary below B, which includes A, exhibits G-IRS. Finally, any boundary DMU between B and E, including D, exhibits G-SCRS.

The above characterisation of GRS can be transferred from our two examples to any technology T, which satisfies assumptions (A1) and (A2). This will be dealt with in the sections below.

To conclude this preliminary outline, two comments should be made. First, if technology T is convex, the first three types of GRS are exhaustive. (This formally follows from Theorem 4.) Therefore, G-SCRS can occur only in a non-convex technology.

Second, if technology T is convex, the average productivity of an efficient DMU would gradually improve as it moves towards its MPSS (see Section 11). If T is not convex, as in Figure 3, this observation is generally not valid. For example, if DMU K reduces its input to that of DMU F, which is on the way to the MPSS unit E, its average productivity will decrease. This is hardly surprising because the types of GRS are defined as indicators of the direction in which the *global*

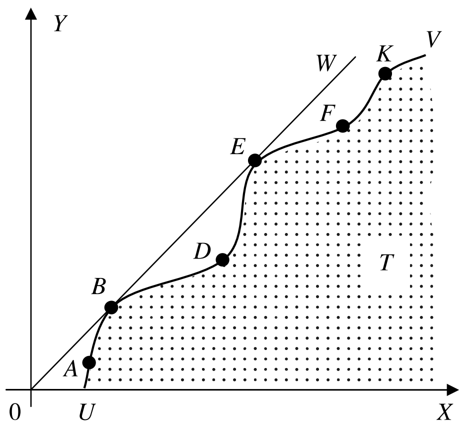

Figure 5. A smooth analogue of the FDH technology.

maximum of the average productivity is achieved, and local improvements in this direction in a non-convex technology are not guaranteed.

6. Global Returns to Scale: The General Case

Consider any production technology T, which satisfies assumptions (A1) and (A2) and, generally, has multiple inputs and outputs, and assume that DMU_0 is efficient in T. By maximising the ratio γ/δ as long as $(\delta X_0, \gamma Y_0)$ remains a member of T, the maximum average productivity for the given mixes X_0 and Y_0 can be identified. This, of course, leads to model (4), in which the reciprocal ratio δ/γ is minimised. Therefore, the DMUs that have the maximum average productivity are the DMUs that operate at MPSS.

The above suggests that, as in Section 5, the type of GRS exhibited by DMU_0 can be identified by comparing the size of MPSS units having the structure $(\delta X_0, \gamma Y_0)$ with the size of DMU_0. This will identify the direction(s) towards the global maximum of the average productivity and lead to the four possible types of GRS.

Due to the fact that assumptions (A1) and (A2) about T are extremely weak, we have to overcome certain difficulties with the attainability of MPSS, as illustrated by Figures 1 and 6. In Figure 1 there is no unit, which operates at MPSS, although the average productivity is decreasing as the efficient DMUs become larger, and therefore GRS should be classed as decreasing. In Figure 6, the only unit operating at MPSS is B, which is larger than DMU A. At the same time the average productivity of the efficient units close to the origin approaches that of DMU B, although never attains it. (The straight line through the origin and point B is tangent to the boundary of the PPS at the origin.) It would, therefore, be counter-intuitive to class A as a G-DRS unit because, by significantly reducing the scale of its operations, DMU A could achieve the average productivity infinitely close to that of B. Therefore, A should be classed as exhibiting G-SCRS.

Definition 3. DMU $(\hat{\delta} X_0, \hat{\gamma} Y_0) \in T$ is called a scale reference unit (SRU) of DMU_0 if $\delta = \hat{\delta}$ and $\gamma = \hat{\gamma}$ is an optimal solution to program (4), that is $\hat{\delta}/\hat{\gamma} = m^*$. Further, for any $\varepsilon \geq 0$, DMU $(\delta_\varepsilon X_0, \gamma_\varepsilon Y_0) \in T$, where $\delta_\varepsilon, \gamma_\varepsilon > 0$, is called an ε-SRU of DMU_0 if $\delta_\varepsilon/\gamma_\varepsilon \leq m^* + \varepsilon$.

Since m^* is the infimum in (4), an ε-SRU of DMU_0 exists for any $\varepsilon > 0$. If ε is small, the ratio $\delta_\varepsilon/\gamma_\varepsilon$ is very close to m^* and the ε-SRU “almost” represents the lowest ratio m^* between the input mix X_0 and output mix Y_0 available in technology T. Obviously, any SRU of DMU_0 is its ε-SRU with $\varepsilon = 0$. A SRU of DMU_0 exists if and only if the infimum m^* in program (4) is attained.

LEMMA 2. *Assume that* DMU_0 *is efficient in* T. *Suppose* $(\delta X_0, \gamma Y_0) \in T$ *for some* $\delta > 0$ *and* $\gamma > 0$ *such that* $\delta/\gamma < 1$. (*In this case* DMU_0 *obviously does not operate at MPSS.*) *Then either both* δ *and* γ *are greater than 1 or both are smaller than 1.*

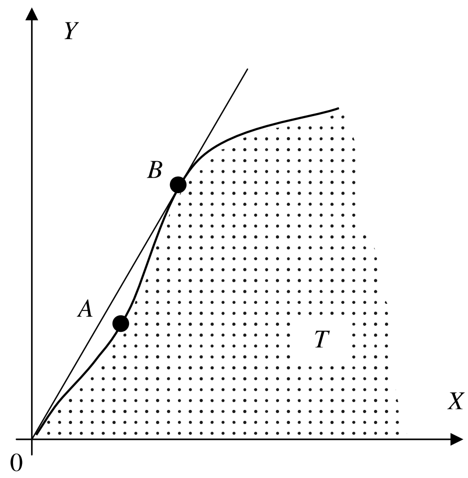

Figure 6. A case of S-CRS with only one SRU.

Suppose that an efficient DMU_0 does not operate at MPSS, that is $m^* < 1$. If we take any small $\varepsilon > 0$ (such that $m^* + \varepsilon < 1$), then for any ε-SRU $(\delta_\varepsilon X_0, \gamma_\varepsilon Y_0)$ we will have, according to Definition 3, $\delta_\varepsilon / \gamma_\varepsilon < 1$. By Lemma 2, this implies that either $\delta_\varepsilon > 1$ and $\gamma_\varepsilon > 1$, or $\delta_\varepsilon < 1$ and $\gamma_\varepsilon < 1$ (but not, e.g., $\delta_\varepsilon < 1$ and $\gamma_\varepsilon > 1$). This means that, for such an ε, any ε-SRU is either *larger* than DMU_0 (if $\delta_\varepsilon > 1$ and $\gamma_\varepsilon > 1$) or *smaller* than DMU_0 (if $\delta_\varepsilon < 1$ and $\gamma_\varepsilon < 1$). This characterisation of ε-SRUs enables us to define the four types of GRS:

Definition 4. Let DMU_0 be efficient in T. Then it exhibits:

(i) G-CRS if DMU_0 operates at MPSS;

(ii) G-DRS if all of its ε-SRUs are smaller than DMU_0 for all sufficiently small ε, (that is for all $\varepsilon \in (0, \bar{\varepsilon})$, where $\bar{\varepsilon}$ is some positive number);

(iii) G-IRS if all of its ε-SRUs are larger than DMU_0 for all sufficiently small ε;

(iv) G-SCRS if, for any $\varepsilon > 0$, some of its ε-SRUs are smaller, and some larger, than DMU_0, but DMU_0 itself does not operate at MPSS.

As an illustration to the above definition, consider DMU A in Figure 1. For all sufficiently small ε, all ε-SRUs of A are close to the origin and, consequently, smaller than A. Therefore, DMU A exhibits G-DRS. For DMU A in Figure 6, the ε-SRUs smaller than A can be found close to the origin. At the same time, a larger unit B is a SRU and thus also an ε-SRU of A for any $\varepsilon > 0$. Consequently, DMU A exhibits G-SCRS.

It is worth noting that the GRS types introduced by Definition 4 are not related in any way to a particular measure of efficiency (input or output radial, additive, etc.), which may be used in the analysis. Therefore, the GRS classification is invariant with respect to the choice of such a measure as long as the PPS is unchanged.

7. Global Returns to Scale: a Special Case

In this section we identify an important and sufficiently large class of technologies for which the concept of ε-SRUs is not required for the definition of GRS. Namely, assume that the PPS T can be represented as the finite union of sets T_k:

$$T = \bigcup_{k=1,\dots,K} T_k, \tag{5}$$

where, for every k, (1) the set T_k satisfies (A1) and (A2) and is convex, and (2) the corresponding CRS technology C_k generated by T_k is closed. It is easy to see that the set T defined by (5) satisfies (A1) and (A2).

It is easy to see that the above assumptions (1) and (2) always hold if every set T_k is a polyhedral (hence convex) set which satisfies (A1) and (A2). In this case the sets C_k are also polyhedral (and therefore, closed). Obvious examples of this include the PPS of the BCC model and the FDH technology, of which the latter is the finite union of polyhedral sets. Further examples include a range of technologies based on the assumption of selective convexity (Podinovski, 2005).

The following theorem classifies GRS using only the SRUs of DMU_0.

THEOREM 2. *Suppose that* DMU_0 *is efficient in* T *defined by* (5). *Then* DMU_0 *has at least one SRU. Further,* DMU_0 *exhibits*

(i) G-CRS *if and only if* DMU_0 *operates at* MPSS*,*

(ii) G-DRS *if and only if all of its* SRU*s are smaller than* DMU_0*,*

(iii) G-IRS *if and only if all of its* SRU*s are larger than* DMU_0*,*

(iv) G-SCRS *if some of its* SRU*s are smaller, and some larger, than* DMU_0*, but* DMU_0 *itself does not operate at* MPSS.

As noted, the FDH technology can be represented in the form (5). Therefore, Theorem 2 can be used for the classification of GRS of its efficient units. If applied to the efficient DMUs in Figure 3, it produces the same characterisation of GRS as discussed in Section 5. Note, however, that the technology in Figure 6 cannot be represented in the form (5) and Theorem 2, if applied to DMU A, would erroneously class its GRS type as G-IRS.

8. Further Discussion

8.1. Global vs. Local RTS

In this section we compare the traditional concept of RTS with the concept of GRS. The former suggests the type of *marginal* resizing (increasing or decreasing) of the efficient unit, which guarantees improvements in the average productivity of the unit. Conceptually, the unit is said to exhibit IRS if changing (increasing or decreasing) its inputs by a small margin leads to its outputs being changed (increased or, respectively, decreased) in a larger proportion. The DRS type is defined in a similar way. These two definitions are refined below for our specific purposes.

The GRS classification suggests the type of resizing which the efficient unit should implement in order to achieve the *global* maximum of the average productivity. Regarding the resizing problem as an optimisation problem with the aim of increasing the average productivity as a function of the size factor, one can envisage that the difference between the concepts of RTS and GRS is the same as the difference between local search and global optimisation. For example, from the local perspective the function of average productivity may be increasing when the unit marginally increases its size (hence the IRS type), while the global optimum might be in the opposite direction (hence the G-DRS type).

In a practical context, the RTS and GRS classifications address different issues. The RTS classes are indicative of the type of resizing of the unit which should result in *immediate* improvements of its average productivity. The GRS classes indicate the direction of change necessary to achieve the *global* maximum of the average productivity. This is obviously useful in assisting strategic decisions like those involving mergers of units or splitting into smaller firms.

It is worth noting that the concept of GRS is well defined under extremely weak assumptions about the production technology, while the traditional definition of RTS is less universal. For example, the RTS classification of the efficient DMUs in the FDH technology in Figure 3 is unclear.

Consider the case in which the concept of RTS is well defined. For simplicity, we shall limit our discussion to the case of one input and one output, but the same conclusions could be reached in the general case.

As an illustration, refer to Figure 7, where the boundary of the technology T is defined by a differentiable function $Y = Y(X)$.

From the global perspective, the only DMU operating at MPSS is DMU A, which therefore, exhibits G-CRS. All the other DMUs B, C, D, E and F are efficient and larger than A, and hence all exhibit G-DRS.

From the local perspective, the classification of RTS is different. Whether the average productivity of an efficient DMU increases, decreases or remains constant to the first degree of approximation depends on the ratio of the marginal productivity of the unit to its average productivity. The former is equal to slope of the boundary $dY(X)/dX$. The latter is the ratio Y/X, that is the slope of the line passing though the origin and the point representing the unit on the graph. The ratio

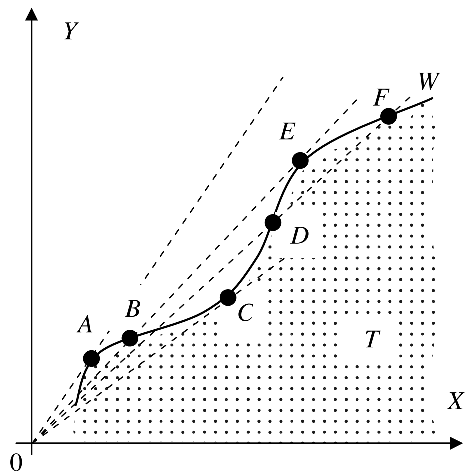

Figure 7. The difference between RTS and GRS.

of these two measures is known as the scale elasticity and is commonly expressed as

$$S(X,Y)=\frac{X}{Y}\frac{dY}{dX}. \tag{6}$$

Consider, e.g., DMU D in Figure 7. As seen in the graph, a marginal increment of its input would result in more than proportional increment of the average productivity, provided the unit continues to operate efficiently. In this case the scale elasticity $S(X,Y)$ is greater than 1, and DMU D is classed as exhibiting IRS. The same type of RTS is exhibited by any other unit on the boundary between C and E. For similar reasons, the scale elasticity at any unit on the boundary between A and C and also to the right of E, is smaller than 1. Consequently, all these units, including B and F, exhibit DRS.

The scale elasticity at three DMUs, A, C and E, is equal to 1, and these units are classed as exhibiting CRS. It is interesting to point out that, although E can be said to operate at the locally most productive scale size, DMU C operates at the locally *least productive scale size*.

To illustrate the relation between RTS and GRS further, rewrite the derivative of the average productivity $Y(X)/X$ in the following form:

$$\frac{d(Y/X)}{dX}=\frac{Y'X-Y}{X^2}=\frac{(X/Y)Y'-1}{(X^2/Y)}=\frac{S(X,Y)-1}{(X^2/Y)} \tag{7}$$

where $S(X,Y)$ is the scale elasticity as defined by (6).

As follows from (7), the set of DMUs, at which the derivative of the function of average productivity is zero, is the set of DMUs exhibiting CRS (where the elasticity is equal to 1). This means that the DMUs exhibiting CRS (more precisely, the inputs X of such DMUs) are stationary points of the function of average productivity. (Similarly, the IRS type corresponds to the areas of strict growth, and DRS, the areas of strict decline, of the average productivity.)

In other words, the condition that an efficient DMU exhibits CRS is *a necessary condition* for this DMU to exhibit G-CRS (and operate at MPSS). As proved in Section 11, this condition is also sufficient if the technology T is convex (although the function of average productivity may not be concave). In a general non-convex technology this condition is not sufficient. The latter is demonstrated by the technology in Figure 7, in which three DMUs, A, C and E, exhibit CRS, but only A operates at MPSS and is classed as G-CRS.

8.2. GRS and Multiproduct Scale Economies

In this section we link our development of GRS to the neoclassical concept of multiproduct scale economies based on the consideration of the cost function (see, e.g., Baumol et al., 1982).

For simplicity, we first refer to the case of a single-product technology depicted in Figure 7, where we shall now assume that X is the total cost of resources making the production of Y possible. The boundary of this technology, represented by the curve AW, defines the cost function $X = c(Y)$. For every Y the corresponding value $c(Y)$ is equal to the minimum cost sufficient for the production of Y. The interior points of the technology T in Figure 7 represent inefficient production units, that is those whose cost of production is higher than the cost function indicates.

For any member (X, Y) of this technology T, the average production cost is defined as $X/Y = c(Y)/Y$. This is obviously reciprocal to the average productivity of the unit defined as Y/X. On the boundary of technology T, if the average productivity is increasing, the inverse ratio, that is the average cost, is decreasing, and vice versa. For example, at the point B in Figure 7, the average productivity is decreasing while the average cost is increasing. Further, both functions of average productivity and average cost have the same sets of stationary points, where their derivatives are equal to zero and the CRS type is observed. In Figure 7, these sets consist of the three points: A, C and E.

The above equivalence implies that the classification of local RTS is identical whether the cost function or average productivity approach is used for the definition of scale economies. This equivalence extends to the case of a multiproduct technology where the scale elasticity does not depend on whether the ray average cost or productivity are used for its definition.

Provided the cost is the only input, the M locus of a multiproduct technology is defined as the set of all product mixes Y, each of which minimises the average cost along the corresponding ray.[4] Since the ray average cost and ray average productivity are mutually reciprocal, the concepts of M locus and MPSS are equivalent. More precisely, the M locus of technology T is equal the set of all output mixes Y such that the unit $(c(Y)), Y)$ operates at MPSS.

We can now restate our main conclusions of this section in terms of the cost function. Namely, the types of GRS are indicative of the direction towards the M locus of the technology, more precisely, its members on the ray under consideration. Mathematically it is possible that there could be a few isolated members of the M locus on a given ray, in which case the function of average costs is obviously not U-shaped. For example, if X in Figure 5 is the production cost, the units B and E represent the M locus (as well as MPSS) and exhibit G-CRS. The other units on the boundary are classed into the remaining three GRS types, which indicate the direction to the M locus.

Further, the local characterisation of RTS is generally not indicative of the relative position of the M locus. For example, as discussed, unit D in Figure 7 exhibits IRS in the local sense and, consequently, its average production cost will decrease if the unit moves towards E, but the M locus is found in the opposite direction, at unit A.

Finally, if the technology T is convex, then, as established in Section 11, the G-SCRS type is impossible and the local RTS classification is consistent with the GRS classification. Therefore, in a convex technology, the local RTS classes are themselves indicative of the direction to the M locus (and MPSS).

9. Testing for Global Returns to Scale

If technology T is convex, the method of testing RTS, as suggested by Färe et al. (1983, 1985), is equally applicable to testing GRS, because the two characterisations of RTS, local and global, are identical (see Theorem 7).

If technology T is not convex, this method should not be applied for testing RTS. As an illustration, refer to the technology T in Figure 7. The corresponding CRS technology C is the closed cone between the ray OA and the X-axis. The NIRS technology H is the area below the consecutively joined segment OA, curved segment AB, segment BE and curved segment EW.

Consider DMU D. Since $E^{\mathrm{i}}_T(D)=1$ and $E^{\mathrm{i}}_C(D)<E^{\mathrm{i}}_H(D)$, unit D is erroneously classed as exhibiting DRS, while, as discussed in Section 8.1, it exhibits IRS. The same applies to the units C and E, which exhibit CRS but are erroneously classed as exhibiting DRS. In fact, any DMU on the efficient boundary above A, which includes B, C, D, E and F, is classed as DRS, which suggests that the method of Färe et al. is, in fact, classifying the units according to the GRS classification, and not RTS.

This is not, however, an entirely accurate observation, as the method does not account for the G-SCRS type and would confuse it with G-IRS. (If the method used the NDRS reference technology instead of the NIRS one, it would confuse the G-SCRS type with G-DRS.) As an illustration consider the FDH technology T in Figure 3. The corresponding CRS technology C is the closed cone between the ray OW and the X-axis. The NIRS technology H is the area below the broken line $OEPKV$. As discussed, DMU D exhibits G-SCRS. However, since $E^{\mathrm{i}}_T(D)=1$ and $E^{\mathrm{i}}_H(D)=E^{\mathrm{i}}_C(D)<1$, DMU D would erroneously be classed as exhibiting G-IRS, if the method of Färe et al. were used for this purpose.

Similarly, DMU A in the technology T in Figure 6 exhibits G-SCRS but would be classed as exhibiting G-IRS, if the same method were applied.

From these examples we can conclude that, *in a non-convex technology*, the method of Färe et al., first, does not test for RTS, and, second, appears to be more consistent with the GRS classification but misses out the G-SCRS type.

Below we suggest a slightly modified approach to testing GRS, equally applicable to convex and non-convex technologies. The difference with the approach of Färe et al. is that, instead of reference technologies C and H, we use G and H, which allows us to distinguish between all four possible types of GRS. The same two reference technologies were used by Kerstens and Vanden Eeckaut (1999) in their modification of the method of Färe et al.[5]

Theorem 3 is formulated in terms of the input radial efficiencies of DMU_0 in technologies T, H and G, where T needs to satisfy only (A1) and (A2). Since DMU_0 is assumed efficient, in each of the four inequalities the term $E^{\mathrm{i}}_T(X_o, Y_o)$ can be substituted by 1. Also, according to Theorem 1, the statement below remains valid if all of the input efficiencies of DMU_0 are substituted by their output analogues.

THEOREM 3. *Suppose DMU_0 is efficient in T. Then it exhibits:*

(i) G-CRS *if and only if* $E^{\mathrm{i}}_G(X_0, Y_0) = E^{\mathrm{i}}_H(X_0, Y_0) = E^{\mathrm{i}}_T(X_0, Y_0)$,

(ii) G-DRS *if and only if* $E^{\mathrm{i}}_G(X_0, Y_0) < E^{\mathrm{i}}_H(X_0, Y_0) \le E^{\mathrm{i}}_T(X_0, Y_0)$,

(iii) G-IRS *if and only if* $E^{\mathrm{i}}_H(X_0, Y_0) < E^{\mathrm{i}}_G(X_0, Y_0) \le E^{\mathrm{i}}_T(X_0, Y_0)$,

(iv) G-SCRS *if and only if* $E^{\mathrm{i}}_G(X_0, Y_0) = E^{\mathrm{i}}_H(X_0, Y_0) < E^{\mathrm{i}}_T(X_0, Y_0)$.

To see that Theorem 3 classes the GRS type of DMU D in Figure 3 correctly, note that the NDRS technology G is the area below the broken line $UAQBW$. Since $E^{\mathrm{i}}_G(D) = E^{\mathrm{i}}_H(D) < 1 = E^{\mathrm{i}}_T(D)$, DMU D is correctly classed as exhibiting G-SCRS.

As another illustration, consider DMU A in Figure 1. Since $E^{\mathrm{i}}_T(A) = E^{\mathrm{i}}_H(A) = 1$ and $E^{\mathrm{i}}_G(A) = 0$, according to Theorem 3, A exhibits G-DRS. Similarly, for DMU B in Figure 2, $E^{\mathrm{i}}_T(B) = E^{\mathrm{i}}_G(B) = 1$ and $E^{\mathrm{i}}_H(B) = 0$. Therefore, B exhibits G-IRS.

We conclude this section by a practical remark concerning the use of Theorem 3 with the FDH technology. Mixed integer non-linear programming models for the assessment of efficiency of DMU_0 in the reference technologies H and G induced by the free disposal hull were developed by Kerstens and Vanden Eeckaut (1999) and Briec et al. (2000). Computationally simpler linear analogues were later suggested by Podinovski (2004a).

10. Global Returns to Scale for Inefficient Units

The issue of MPSS and GRS for inefficient DMUs is, with one notable exception, a matter of definition, and different definitions may, in some cases, lead to different characterisations.

Definition 5. DMU_0 is called *R*-efficient (*R* stands for the radial nature of efficiency), if $E_T^{\text{i}}(X_0, Y_0) = E_T^{\text{o}}(X_0, Y_0) = 1$.

Any efficient DMU is *R*-efficient but the converse is generally not true. As follows from Definition 2, any unit operating at MPSS is *R*-efficient, but not necessarily efficient, in technologies T and C. Consequently, any SRU of a DMU_0 is an *R*-efficient DMU.

Revisiting the proof of Lemma 2, it is easy to see that it remains valid for any DMU_0, which is *R*-efficient. Therefore, the classification of the types of GRS based on this lemma, as given by Definition 4, is equally applicable to any *R*-efficient DMU_0.

Further, as the proofs of Theorems 1, 2 and 3 really use only the fact of *R*-efficiency of DMU_0, and not its efficiency in the sense of Definition 1, their statements remain valid for any *R*-efficient DMU_0. In summary, there is no difference between efficient and *R*-efficient DMUs as far as the definition and testing of GRS are concerned.

If an inefficient DMU_0 is not *R*-efficient, it is possible that it has a SRU $(\delta X_0, \gamma Y_0)$, where $\delta < 1$ and $\gamma > 1$, which is neither larger nor smaller than DMU_0. Consequently, the GRS as an indicator of the direction leading to MPSS is undefined. An example of this type is DMU L in Figure 3, which has two SRUs, B and E, the latter being neither smaller nor larger than L.

Further classification rules can be added to the definition of GRS, similar to those used to classify RTS of inefficient DMUs, in order to expand its applicability to some, or even all, inefficient units. For example, DMU P in Figure 3 could be classed as exhibiting G-DRS because both of its SRUs B and E could be regarded as being smaller than P. Another approach, leading to a generally different characterisation, was used by Färe et al. (1994) and Cooper et al. (2000). According to it, an inefficient DMU is first projected on the efficient frontier of T, and the type of GRS (or RTS, as in Färe et al. and Cooper et al.) is defined at this projection. If this were applied to DMU P, it would be classed as exhibiting G-CRS.

Since the classification of the GRS types of inefficient units is a matter of definition, we are not taking this issue any further.

11. Returns to Scale in a Convex Technology

Suppose that, in addition to assumptions (A1) and (A2), T is a convex set. The aim of this section is twofold. First, we prove that the case of G-SCRS is impossible in T. Second, we show that the types of GRS in technology T are consistent with the classification of local RTS.

11.1. Impossibility of the G-SCRS Type

THEOREM 4. *If technology T is convex, the case of* G-SCRS *is impossible.*

This theorem[6] and the obvious equality $E_C(X_0, Y_0) = \min(E_G^{\text{i}}(X_0, Y_0), E_H^{\text{i}}(X_0, Y_0))$ imply that, in a convex technology, the original method of Färe et al. and its modification stated in Theorem 3 are equivalent.

Further, consider the convex technology of the BCC model. As Banker et al. (1996) shows, in this technology, the original method of Färe et al. is equivalent to the primal (also known as the CCR RTS) and dual (BCC RTS) methods of testing local types of RTS. This has two implications.

First, if according to Theorem 3 the method of Färe et al. classes DMU_0 as exhibiting, say, G-IRS, then DMU_0 also exhibits local IRS, and vice versa. In the same way, G-DRS corresponds to the type of local DRS, and G-CRS to local CRS. In other words, in the BCC model the local and global RTS classes are identical. Below, this result is extended to any closed convex technology.

Second, several methods have been developed for the testing of types of RTS in the BCC model (see, e.g., Banker et al., 2004). Since the local and global RTS classes are identical in the BCC model, all such methods, including the one based on the sign of the dual variable of the convexity constraint, are also testing for the types of GRS.

11.2. Ray Average Productivity

The assumptions (A1) and (A2) are too weak for our further development. In this and the following sections we shall additionally assume that the convex PPS T is a closed set. Some of the results below or their analogues can be formulated without the assumption of closedness, but the overall development becomes excessively technical.

Consider any efficient[7] DMU_0 in the technology T. This DMU becomes a base unit in the development below. Define Δ, the set of all $\delta > 0$ such that, for some $\gamma > 0$, $(\delta X_0, \gamma Y_0) \in T$. Since the PPS T is convex, Δ is also a convex set. In other words, Δ is an interval, which may be bounded or unbounded, or a single-point set.[8] The set of all vectors δX_0, where $\delta > 0$, is the ray in the input space induced by the input vector X_0, and the interval $\{\delta X_0 | \delta \in \Delta\}$ is its subset.

For every $\delta \in \Delta$, introduce the function $\bar{\gamma}(\delta)$ as

$$\bar{\gamma}(\delta) = \max\{\gamma \,|\, (\delta X_0, \gamma Y_0) \in T\}. \tag{8}$$

Due to assumption (A1), $\bar{\gamma}(\delta)$ is finite for every $\delta \in \Delta$. Since T is closed, $\bar{\gamma}(\delta)$ is attained and represents the maximum proportion of output Y_0 obtainable in technology T for the vector of inputs δX_0. In particular, since DMU_0 is efficient,

$$\bar{\gamma}(1) = 1. \tag{9}$$

LEMMA 3. *The function $\bar{\gamma}(\delta)$ is concave and continuous on Δ.*

It is worth noting that the unit

$$(\delta X_0, \bar{\gamma}(\delta) Y_0). \tag{10}$$

where $\delta \in \Delta$, is not necessarily efficient in technology T, although, by definition of $\bar{\gamma}(\delta)$, its output radial efficiency is always equal to one. For example, suppose that, in the FDH technology in Figure 3, we have selected unit A as the base DMU_0, so that $\delta = 1$ corresponds to A and $\Delta = [1, +\infty)$. Then, of all the units in the form (10), only six of them: A, B, D, E, F and K, are efficient, while the rest are not.

Definition 6. The ray average productivity (RAP) of the unit $(\delta X_0, \bar{\gamma}(\delta) Y_0)$, where $\delta \in \Delta$, is

$$\varphi(\delta) = \frac{\bar{\gamma}(\delta)}{\delta}. \tag{11}$$

The definition of RAP relates the quantity $\bar{\gamma}(\delta)$ of the fixed output bundle Y_0 to the quantity δ of the fixed input bundle X_0. This generalises the notion of average productivity discussed in Section 8.1 for the simple case with one input and one output. Since, obviously, $\varphi(1) = 1$, the RAP $\varphi(\delta)$ is the factor by which the average productivity of the unit (10) differs from the average productivity of DMU_0.

LEMMA 4. *The function $\varphi(\delta)$ is continuous and quasiconcave on Δ, that is*

$$\varphi(\delta_2) \geq \min(\varphi(\delta_1), \varphi(\delta_3)) \tag{12}$$

for any δ_1, δ_2 and $\delta_3 \in \Delta$ such that $\delta_1 < \delta_2 < \delta_3$. Further, if $\varphi(\delta_1) \neq \varphi(\delta_3)$, the inequality in (12) *is strict.*

It is worth emphasising that, although the technology T is assumed convex and the function $\bar{\gamma}(\delta)$ is concave on Δ, the ray average productivity $\varphi(\delta)$ is generally not concave. For example, refer to the technology shown in Figure 1 and define $X_0 = 1$. Then $Y_0 = 1$, and for every efficient unit (10) we have $\varphi(\delta) = \sqrt{\delta}/\delta = 1/\sqrt{\delta}$, which is not a concave function.

In the above development, DMU_0 was used as the base unit to define the interval Δ and functions $\bar{\gamma}(\delta)$ and $\varphi(\delta)$. Suppose that, instead of DMU_0, another unit

$$(\tilde{X}, \tilde{Y}) = (\tilde{\delta} X_0, \bar{\gamma}(\tilde{\delta}) Y_0) \tag{13}$$

from the same ray is chosen as the base unit. This redefines the interval Δ and the functions $\bar{\gamma}(\delta)$ and $\varphi(\delta)$. In particular, $\delta = 1$ will now correspond to DMU $(\tilde{X}, \tilde{Y})$. Further, the RAP of any unit on the ray changes by the positive factor $\tilde{\delta}/\bar{\gamma}(\tilde{\delta})$, where $\tilde{\delta}$ and $\bar{\gamma}(\tilde{\delta})$ are as in (13). Since this factor is constant, the general behaviour of the function of RAP remains unchanged on the ray. More specifically, the RAP at any DMU (X, Y) is increasing or decreasing irrespective of the base unit (13), and the sign of the derivative $\varphi(\delta)$ is also invariant in this respect.[9] In our development below this means that the characterisation of GRS and local scale economies on any ray does not depend on the choice of the base unit DMU_0.

11.3. The Structure of Ray Average Productivity

Let ϕ^* be the supremum[10] of $\varphi(\delta)$ on the interval Δ. Define Δ^*, the set of all $\delta \in \Delta$ such that $\varphi(\delta) = \varphi^*$. If the supremum ϕ^* is not attained, $\Delta^* = \varnothing$

Lemma 5. *Provided $\Delta^* \neq \varnothing$, the set Δ^* is closed and convex.*

According to Lemma 5, the set Δ^* may be a single-point set or a sub-interval[11] of Δ. Consequently, the interval Δ can be represented as the union of three intervals:

$$\Delta = \Delta^+ \cup \Delta^* \cup \Delta^-, \tag{14}$$

where Δ^+ is the set of all $\delta \in \Delta$ such that $\delta < \delta'$ for any $\delta' \in \Delta^*$. Similarly, Δ^- is the set of all $\delta \in \Delta$ such that $\delta > \delta'$ for any $\delta' \in \Delta^*$.

In representation (14), each of the three sub-intervals Δ^+, Δ^* and Δ^-, but obviously not all of them, can be empty sets.[12] The only impossible nontrivial case is where the set Δ^* is empty but both sets Δ^+ and Δ^- are not. This fact is established in Theorem 5.

Theorem 5. *The ray average productivity $\varphi(\delta)$ is strictly increasing on Δ^+ and strictly decreasing on Δ^- (provided these sets are not empty). If $\Delta^+ \neq \varnothing$ and $\Delta^- \neq \varnothing$, then $\Delta^* \neq \varnothing$.*

As an illustration to Theorem 6, refer to the technology T whose efficient boundary is defined by the equation $Y = \ln(X+1)$, as shown in Figure 8. To be specific, let point A be DMU_0. In this case $\Delta = (0, +\infty)$ and the average productivity $\varphi(\delta)$ is strictly decreasing on the entire interval Δ, so that $\Delta^- = \Delta$. The maximum (supremum) of average productivity $\varphi(\delta)$ is not attained, and $\Delta^* = \Delta^+ = \varnothing$. Obviously, for all sufficiently small $\varepsilon > 0$, the ε-SRUs of any efficient DMU_0 in T will be located close to the origin.

11.4. GRS, RTS and Scale Economies

As an immediate implication of Theorem 5 we see that, in a closed convex technology T, the notion of GRS is consistent with the notion of scale economies. In particular, scale economies are present on the interval Δ^+ where all units belong to the G-IRS type. Similarly, diseconomies of scale are observed on the interval Δ^-, where all units exhibit G-DRS. Finally, economies of scale are exhausted on the interval Δ^*, where the MPSS is achieved and G-CRS are observed.

We shall now investigate a more subtle correspondence between the GRS and RTS classifications. As we were able to see, in a closed convex technology, the former is equivalent to the notion of scale economies. However, as highlighted by

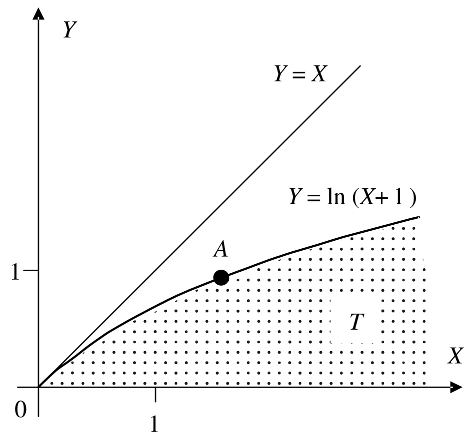

Figure 8. Unattained finite maximum productivity at the origin.

Baumol et al. (1982), scale economies do not necessarily imply IRS, as well as scale diseconomies do not imply DRS.

The discrepancy between the two notions arises because the strict monotonicity of the function of ray average productivity $\varphi(\delta)$ does not in principle preclude its derivative from taking on a zero value. The corresponding DMU would exhibit CRS and economies of scale at the same time. Lemma 6 shows that, in a convex technology, this cannot happen.

Lemma 6. *Let $\tilde{\delta}$ be any interior point of the set Δ. Assume that the function $\bar{\gamma}(\delta)$ and, therefore, $\varphi(\delta)$ are differentiable in some neighbourhood of $\tilde{\delta}$. If $\varphi(\delta)$ is strictly increasing at $\tilde{\delta}$, then $\varphi'(\tilde{\delta}) > 0$. If $\varphi(\delta)$ is strictly decreasing at $\tilde{\delta}$, then $\varphi'(\tilde{\delta}) < 0$.*

We now need to translate the above statement to the language of RTS. Suppose that the boundary of the PPS T, on which the efficient DMU_0 is located, is described, at least in some neighbourhood of DMU_0, by the equation $F(X, Y) = 0$, where $X \in R^m$ and $Y \in R^s$. Since all units (10) are located on the boundary of T, we have

$$F(\delta X_0, \bar{\gamma}(\delta) Y_0) = 0 \tag{15}$$

for all $\delta \in \Delta$ which are sufficiently close to $\delta = 1$ (the latter corresponds to DMU_0). In particular, for DMU_0 we have $F(X_0, Y_0) = 0$.

We should now assume that the standard assumptions of the implicit function theorem[13] are satisfied for equation (15) at its solution $(\delta, \bar{\gamma}) = (1, 1)$, or simply at DMU_0. These assumptions are a standard requirement for the definition of the types of local RTS using the notion of scale elasticity[14]. Eventually, this leads to the main result of this section, Theorem 7, which establishes the equivalence of the GRS and RTS classifications.

The type of RTS exhibited by DMU_0 depends on the scale elasticity (see, e.g., Panzar and Willig, 1977)

$$S(X_0, Y_0) = -\frac{\langle X_0, \nabla_X F(X_0, Y_0)\rangle}{\langle Y_0, \nabla_Y F(X_0, Y_0)\rangle}, \tag{16}$$

where $\nabla_X F(X_0, Y_0)$ and $\nabla_Y F(X_0, Y_0)$ are the partial gradients of the function $F(X, Y)$ by, respectively, X and Y at DMU_0, and $\langle,\rangle$ denotes the scalar product.

LEMMA 7. *Assume that the production correspondence* (15) *satisfies the assumptions of the implicit function theorem at* DMU_0. *Then*

$$S(X_0, Y_0) = \varphi'(\delta)\big|_{\delta=1} + 1. \tag{17}$$

Lemma 7 links the scale elasticity to the derivative of the function of ray average productivity at DMU_0 and explains the rationale behind the following definition of RTS.

Definition 7. DMU_0 is said to exhibit IRS, CRS and DRS if, respectively, $S(X_0, Y_0) > 1$, $S(X_0, Y_0) = 1$ and $S(X_0, Y_0) < 1$.

Taking into account Lemma 7, the definition of RTS is equivalent to the following classification: DMU_0 exhibits IRS, CRS and DRS if, respectively, $\varphi'(1) > 0$, $\varphi'(1) = 0$ and $\varphi'(1) < 0$. This is consistent with the single-input and single-output case discussed in Section 8.1.

Combining Lemmas 6 and 7, and taking into account representation (14), we can equate the types of GRS to the corresponding types of RTS.

THEOREM 7. *Assume that the production correspondence* (15) *satisfies the assumptions of the implicit function theorem at* DMU_0. *Then* DMU_0 *exhibits* G-IRS, G-DRS *or* G-CRS *if and only if it exhibits, respectively,* IRS, DRS *or* CRS.

If the implicit function theorem does not apply to DMU_0, the scale elasticity and based on it classification of RTS are undefined, and Theorem 7 becomes invalid. However, as shown at the beginning of this section, the GRS classes remain consistent with the notion of economies of scale, as they always are in a closed convex technology.

An example of this situation is the polyhedral technology of the BCC DEA model. Since the efficient boundary of such a technology is not smooth at the edges formed by the intersection of hyperplanes, the scale elasticity, and therefore, type of RTS, is undefined at any DMU on the edge. For example, in technology T depicted in Figure 9, the scale elasticity and, therefore, local types of RTS based on it are undefined at units A, B and C. In practice, however, the units A and B would be classed as exhibiting IRS, and unit C as exhibiting CRS. This is consistent with the fact that units A and B exhibit scale economies, and at unit C the scale economies are exhausted. This extension of the RTS classification is entirely consistent with the notion of GRS because the latter is equivalent to the notion of scale economies.

From the practical perspective, Theorem 7 implies that, if DMU_0 gradually changes the scale of its operations towards its MPSS and produces the maximum output for every intermediate vector of inputs throughout this process, its average productivity strictly increases until the MPSS is reached.

12. Conclusion

In this paper a concept of global returns to scale was developed as an indicator of the direction in which an efficient DMU should change the scale of its operations in order to achieve the global maximum average productivity without changing its input and output proportions. The GRS classification includes constant, increasing, decreasing and sub-constant types of global RTS. In a convex technology, the first three of these coincide with the traditional local RTS types, while the sub-constant type is impossible. In a non-convex technology the GRS and RTS classifications are generally different.

It was shown that, if the technology is not convex, the well-known method of Färe et al. for testing RTS is, in fact, testing for GRS although not accounting for the sub-constant type of GRS. In the convex case this difference is not observed, as the concepts of GRS and RTS are identical. A simple modification of this method was suggested, which made it suitable for the identification of all four types of GRS.

It was demonstrated that the concepts of efficiency, MPSS and GRS are so fundamental that can be correctly defined and tested under an extremely weak assumption about the production technology. In particular, the production technology is not required to be convex, smooth, closed or exhibit any kind of

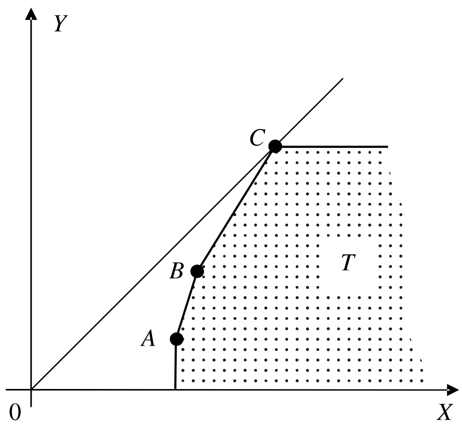

Figure 9. The polyhedral technology of the BCC model.

disposability of inputs and outputs. This contrasts with the traditional concept of RTS, which is based on the notion of scale elasticity and requires the production correspondence to be differentiable at the unit under consideration.

From a practical point of view, the GRS classes should be useful in strategic decision situations, as they indicate the global maximum of the average productivity. In contrast, the traditional local classification of RTS does not, unless the technology is convex, indicate the direction to the MPSS. The RTS classification should be useful in assisting decisions concerning relatively minor changes in the scale of operations.

Appendix: Proofs

Additional Notation for the Appendix

To simplify notation, we denote $E_T^{\mathrm{i}} = E_T^{\mathrm{i}}(X_0, Y_0)$ and define E_H^{i}, E_H^{o}, E_G^{i}, E_G^{o} and E_C in a similar way.

Proof of Lemma 1. To prove (A1), assume $X \neq 0$ and there exists a sequence $\{(X, Y_k)\} \subset T$ such that $\|Y_k\| \to \infty$ as $k \to \infty$. Define $X_{(k)} = X/\|Y_k\|$ and $Y_{(k)} = Y/\|Y_k\|$. Clearly, $(X_{(k)}, Y_{(k)}) \in \bar{C}$ and $\|Y_{(k)}\| = 1$ for every k. Since all $Y_{(k)}$ are on the unit sphere, without loss of generality, $\{(X_k, Y_k)\} \to (0, \hat{Y})$, where $\|\hat{Y}\| = 1$. Since $\bar{C}$ is closed, $(0, \hat{Y}) \in \bar{C}$, which is impossible due to (1). Finally, if the infimum in (A2) is zero, it is attained in $\bar{C}$ at $X = 0$, and (1) is false. ■

Proof of Theorem 1. Part (i) follows from the fact that $(\theta X_0, Y_0) \in C$ if and only if $(X_0, (1/\theta)Y_0) \in C$, for any $\theta > 0$. (ii) For any $\varepsilon > 0$, there exist $\theta \in (0, E_H^{\mathrm{i}} + \varepsilon)$, $(X', Y') \in T$ and $\alpha \in (0, 1]$ such that $(\theta X_0, Y_0) = \alpha(X', Y')$. Then $\theta/\alpha \geq 1$, as otherwise (X', Y') dominates (X_0, Y_0). Then $(X_0, (1/\theta)Y_0) = (\alpha/\theta)(X', Y') \in H$ and $E_H^{\mathrm{o}} \leq \theta \leq E_H^{\mathrm{i}} + \varepsilon$. Since ε is arbitrarily small, $E_H^{\mathrm{o}} \leq E_H^{\mathrm{i}}$.

Similarly, for any $\varepsilon > 0$, there exist $\theta \leq \min(E_H^{\mathrm{o}} + \varepsilon, 1)$, $(X', Y') \in T$ and $\alpha \in (0, 1]$ such that $(X_0, (1/\theta)Y_0) = \alpha(X', Y')$. Since $\theta\alpha \leq 1$, $(\theta X_0, Y_0) = \theta\alpha(X', Y') \in H$. Then $E_H^{\mathrm{i}} \leq \theta \leq E_H^{\mathrm{o}} + \varepsilon$. Since ε is arbitrarily small, $E_H^{\mathrm{i}} \leq E_H^{\mathrm{o}}$. Thus $E_H^{\mathrm{i}} = E_H^{\mathrm{o}}$.

The equality $E_G^{\mathrm{i}} = E_G^{\mathrm{o}}$ is proved in a similar way. ■

Proof of Lemma 2. Since $\delta/\gamma < 1$, the case $\delta \geq 1$, $\gamma \leq 1$ is impossible. If $\delta \leq 1$ and $\gamma \geq 1$, DMU_0 is dominated and thus inefficient. ■

Proof of Theorem 2. Without loss of generality, for any $\varepsilon > 0$ such that $m^* + \varepsilon < 1$ there exists an ε-SRU $(\delta_\varepsilon X_0, \gamma_\varepsilon Y_0) \in T_1$ smaller than DMU_0. Let us prove that there also exists a SRU of DMU_0 in T_1. Since C_1 is closed,

$$m^* = \inf\{\delta/\gamma \,|\, (\delta X_0, \gamma Y_0) \in T_1, \delta, \gamma > 0\} = \min\{\delta \,|\, (\delta X_0, Y_0) \in C_1, \delta \geq 0\}$$

and m^* is attained at some δ^*. By definition of C_1, $(\delta^* X_0, Y_0) = \alpha^*(X^*, Y^*)$, where $(X^*, Y^*) \in T_1$ and $\alpha^* \geq 0$. Since $Y_0 \neq 0$, we have $\alpha > 0$. Define $\hat{\delta} = \delta^*/\alpha^*$ and $\hat{\gamma} = 1/\alpha^*$. Then m^* is attained at $(\hat{\delta} X_0, \hat{\gamma} Y_0) \in T_1$, which is a SRU of DMU_0.

Let us prove that this SRU is smaller than DMU_0. Assume that this is not true, that is $\hat{\delta} > 1$, and let $\lambda = (\hat{\gamma} - 1)/(\hat{\gamma} - \varepsilon)$. Then $\lambda \in (0, 1)$ and

$$\lambda(\delta_\varepsilon X_0, \gamma_\varepsilon Y_0) + (1 - \lambda)(\hat{\delta} X_0, \hat{\gamma} Y_0) = (\delta' X_0, Y_0) \in T,$$

where $\delta' = \lambda\delta_\varepsilon + (1 - \lambda)\hat{\delta} < 1$, which contradicts the efficiency of DMU_0. Thus $\hat{\delta} < 1$.

Similarly, if for any $\varepsilon > 0$ there exists an ε-SRU larger than DMU_0, then there exists a SRU larger than DMU_0. Applying these two observations to the conditions for G-DRS, G-IRS and G-SCRS in Definition 4, we obtain the statement of the theorem. ■

Proof of Theorem 3. (i) If $E^{\mathrm{i}}_G = E^{\mathrm{i}}_H = E^{\mathrm{i}}_T$, then, since $E^{\mathrm{i}}_T = 1$, $m^* = E_C = \min(E^{\mathrm{i}}_G, E^{\mathrm{i}}_H) = 1$, and DMU_0 exhibits G-CRS.

(ii) If $m^* = E_C = E^{\mathrm{i}}_G < E^{\mathrm{i}}_H \le 1$, let $m^* + \bar{\varepsilon} < E^{\mathrm{i}}_H$. Then for any $\varepsilon \in (0, \bar{\varepsilon})$, $\delta_\varepsilon/\gamma_\varepsilon < 1$. If $\delta_\varepsilon \ge 1$, then $\gamma_\varepsilon > 1$ and $((\delta_\varepsilon/\gamma_\varepsilon)X_0, Y_0) \in H$. Then $E^{\mathrm{i}}_H \le \delta_\varepsilon/\gamma_\varepsilon < E^{\mathrm{i}}_H$. Contradiction. Thus $\delta_\varepsilon < 1$ and DMU_0 exhibits G-DRS.

(iii) If $m^* = E_C = E^{\mathrm{i}}_H < E^{\mathrm{i}}_G \le 1$, let $m^* + \bar{\varepsilon} < E^{\mathrm{i}}_G$. Then for any $\varepsilon \in (0, \bar{\varepsilon})$, $\delta_\varepsilon/\gamma_\varepsilon < 1$. If $\delta_\varepsilon \le 1$, then $(X_0, (\gamma_\varepsilon/\delta_\varepsilon)Y_0) \in G$ and $E^{\mathrm{i}}_G \le \delta_\varepsilon/\gamma_\varepsilon < E^{\mathrm{i}}_G$. Contradiction. Thus $\delta_\varepsilon > 1$ and DMU_0 exhibits G-IRS.

(iv) If $m^* = E^{\mathrm{i}}_G = E^{\mathrm{i}}_H < 1$, let $m^* + \bar{\varepsilon} < 1$ and $\varepsilon \in (0, \bar{\varepsilon})$. Since $E^{\mathrm{i}}_H = m^*$, there exist $\theta \le m^* + \varepsilon$ and $(X', Y') \in T$ such that $(\theta X_0, Y_0) = \alpha(X', Y')$. Since $\theta < E^{\mathrm{i}}_T$ and $Y_0 \ne 0$, $\alpha \in (0, 1)$. Then $(X', Y') = ((\theta/\alpha)X_0, (1/\alpha)Y_0)$ is the ε-SRU larger than DMU_0.

Similarly, since $E^{\mathrm{i}}_G = m^*$, there exist $\theta \le m^* + \varepsilon$ and $(X', Y') \in T$ such that $(\theta X_0, Y_0) = \alpha(X', Y')$. Since $\theta < E^{\mathrm{i}}_T$, $\alpha > 1$, thus $1/\alpha < 1$. Then $((\theta/\alpha)X_0, (1/\alpha)Y_0)$ is the ε-SRU smaller than DMU_0. Therefore, DMU_0 exhibits G-SCRS. ■

Proof of Theorem 4. Assume DMU_0 exhibits G-SCRS. Then, by Theorem 3, $(\theta' X_0, Y_0) = \alpha(X', Y')$ and $(\theta'' X_0, Y_0) = \beta(X'', Y'')$, for some (X', Y'), $(X'', Y'') \in T$, θ', $\theta'' < E^{\mathrm{i}}_T$ and $\alpha < 1 < \beta$. Let $\lambda = (\beta - 1)/(\beta - \alpha)$. Then $\lambda(X', Y') + (1 - \lambda)(X'', Y'') = (\hat{\theta} X_0, Y_0) \in T$, where $\hat{\theta} < E^{\mathrm{i}}_T$, which contradicts the definition of E^{i}_T. ■

Proof of Lemma 3. This follows immediately from Lemma 10 proved below. ■

Proof of Lemma 4. The continuity of $\varphi(\delta)$ follows from the continuity of $\bar{\gamma}(\delta)$. To simplify notation, for any $\delta_k \in \Delta$ and $\lambda \in [0, 1]$ define

$$\gamma_k = \bar{\gamma}(\delta_k), \qquad \delta^\lambda = \lambda\delta_1 + (1 - \lambda)\delta_2, \qquad \gamma^\lambda = \lambda\gamma_1 + (1 - \lambda)\gamma_2. \tag{A.1}$$

To be specific, assume that $\varphi(\delta_3) \ge \varphi(\delta_1)$, that is $\gamma_3/\delta_3 \ge \gamma_1/\delta_1$. (The proof in the case $\varphi(\delta_3) < \varphi(\delta_1)$ is similar.) Using notation (A.1), we have $\gamma_2 = \bar{\gamma}(\lambda\delta_1 + (1 - \lambda)\delta_3)$, where $\lambda = (\delta_3 - \delta_2)/(\delta_3 - \delta_1)$. By Lemma 3, $\bar{\gamma}(\delta)$ is concave and, therefore,

$$\gamma_2 \ge \frac{\delta_3 - \delta_2}{\delta_3 - \delta_1}\gamma_1 + \frac{\delta_2 - \delta_1}{\delta_3 - \delta_1}\gamma_3. \tag{A.2}$$

Dividing both parts by δ_2, subtracting γ_1/δ_1 from them and rearranging the terms, we obtain

$$\frac{\gamma_2}{\delta_2} - \frac{\gamma_1}{\delta_1} \ge \frac{\delta_2 - \delta_1}{\delta_1\delta_2(\delta_3 - \delta_1)}(\delta_1\gamma_3 - \delta_3\gamma_1). \tag{A.3}$$

Since $\gamma_3/\delta_3 \geq \gamma_1/\delta_1$, we have $\varphi(\delta_2) \geq \varphi(\delta_1)$, and (12) follows. If $\varphi(\delta_1) \neq \varphi(\delta_3)$, the inequality in (A.3) is obviously strict. ■

Proof of Lemma 5. Since $\varphi(\delta)$ is a concave function, the set Δ^* is convex. Since $\varphi(\delta)$ is continuous, the set Δ^* is closed. ■

Proof of Theorem 5. Choose any $\delta_3 \in \Delta^*$ and $\delta_1, \delta_2 \in \Delta^+$ such that $\delta_1 < \delta_2 < \delta_3$. Since $\varphi(\delta_1) < \varphi(\delta_3)$, by Lemma 4, $\varphi(\delta_2) > \varphi(\delta_1)$. Therefore, $\varphi(\delta)$ is strictly increasing on Δ^+. The case of Δ^- is considered in the same way.

Suppose that $\Delta^+ \neq \emptyset$ and $\Delta^- \neq \emptyset$. Choose any $\delta_1 \in \Delta^+$ and $\delta_2 \in \Delta^-$. Since $\bar{\gamma}(\delta)$ is continuous on $[\delta_1, \delta_2]$, it attains its maximum at some $\delta^* \in [\delta_1, \delta_2]$. The point δ^* cannot be interior to Δ^+ and Δ^-. This is only possible if $\delta^* \in \Delta^*$, and, consequently, $\Delta^* \neq \emptyset$. ■

Proof of Lemma 6. Suppose $\varphi(\delta)$ is strictly increasing at $\tilde{\delta}$. By definition, there exists a neighbourhood $(\tilde{\delta} - \varepsilon, \tilde{\delta} + \varepsilon) \subset \Delta$ of $\tilde{\delta}$, where $\varepsilon > 0$, such that, for any $\delta_1 \in (\tilde{\delta} - \varepsilon, \tilde{\delta})$ and $\delta_2 \in (\tilde{\delta}, \tilde{\delta} + \varepsilon)$,

$$\varphi(\delta_1) < \varphi(\tilde{\delta}) < \varphi(\delta_2). \tag{A.4}$$

Without a loss of generality, $\bar{\gamma}(\delta)$ and $\varphi(\delta)$ are differentiable on $(\tilde{\delta} - \varepsilon, \tilde{\delta} + \varepsilon)$. From (11),

$$\varphi'(\delta) = \frac{\delta \bar{\gamma}'(\delta) - \bar{\gamma}(\delta)}{\delta^2}. \tag{A.5}$$

By Lemma 11 proved below, the numerator $\delta \bar{\gamma}'(\delta) - \bar{\gamma}(\delta)$ of (A.5) is not increasing on the interval $(\tilde{\delta} - \varepsilon, \tilde{\delta} + \varepsilon)$. Since $\varphi(\delta)$ is strictly increasing at $\tilde{\delta}$, $\varphi'(\tilde{\delta}) \geq 0$. Suppose that $\varphi'(\tilde{\delta}) = 0$. Then $\varphi'(\delta) \leq 0$ for all $\delta \in [\tilde{\delta}, \tilde{\delta} + \varepsilon)$, which contradicts (A.4). Therefore, $\varphi'(\tilde{\delta}) > 0$.

The case where $\varphi(\delta)$ is strictly decreasing at $\tilde{\delta}$ is proved in a similar way. ■

Proof of Lemma 7. By the implicit function theorem, $\bar{\gamma}(\delta)$ is differentiable on some interval $\delta \in (1 - \varepsilon, 1 + \varepsilon) \subset \Delta$, where $\varepsilon > 0$, and

$$\bar{\gamma}'(\delta) = -\frac{\partial F / \partial \delta}{\partial F / \partial \bar{\gamma}} = -\frac{\langle X_0, \nabla_X F(\delta X_0, \bar{\gamma} Y_0) \rangle}{\langle Y_0, \nabla_Y F(\delta X_0, \bar{\gamma} Y_0) \rangle}.$$

Taking into account (9) and (16), we have $S(X_0, Y_0) = \bar{\gamma}'(1)$, and (17) follows from (A.5). ■

Proof of Theorem 7. Under the conditions of the implicit function theorem, $\delta = 1$ is interior to the interval Δ. Suppose DMU_0 exhibits G-IRS. Then $1 \in \Delta^+$. By Theorem 5 and Lemma 6, $\varphi'(1) > 0$, which implies that DMU_0 exhibits IRS. Conversely, if DMU_0 exhibits IRS, by Definition 7 and Lemma 7, $\varphi'(1) > 0$, which is only possible if $1 \in \Delta^+$. Then DMU_0 exhibits G-IRS. Similarly, the G-DRS class coincides with the DRS class. If DMU_0 exhibits G-CRS, it cannot exhibit IRS or DRS, as this equates to G-IRS and G-DRS. Therefore, DMU_0 exhibits CRS, and the theorem follows. ■

LEMMA 8. *Suppose that the PPS T is closed, convex and satisfies assumption* (A1). *Define*

$$\Delta^0 = \{\delta \geq 0 | \exists \gamma : (\delta X_0, \gamma Y_0) \in T\} \tag{A.6}$$

(Note that $\Delta = \Delta^0/\{0\}$.) Consider any bounded subset $\tilde{\Delta}$ of Δ^0. (This means that there exists an $R > 0$ such that $\delta < R$ for any $\delta \in \Delta$.) Then there exists an $M > 0$ such that $\bar{\gamma}(\delta) < M$ for every $\delta \in \tilde{\Delta}$, that is the function $\bar{\gamma}(\delta)$ is bounded above on $\tilde{\Delta}$.

Proof of Lemma 8. Assume that the statement is false. Then there exists an infinite sequence of pairs $(\delta_k, \gamma_k), k = 1, 2, \ldots$, such that, for every k, $\delta_k \in \tilde{\Delta}$, $(\delta_k X_0, \gamma_k Y_0) \in T$ and $\gamma_k \to +\infty$ when $k \to +\infty$. Due to the boundness, the sequence $\{\delta_k\}$ can be assumed converging to some $\tilde{\delta}$ and the corresponding $\{\gamma_k\}$ strictly increasing. By taking an infinite subsequence of $\{\delta_k\}$ to one side of $\tilde{\delta}$, we can assume, without loss of generality, that $\{\delta_k\}$ is strictly increasing.

For every $k \geq 3$, define l_k, the line segment joining (δ_1, γ_1) with (δ_k, γ_k), and γ_k^* such that $(\delta_2, \gamma_k^*) \in l_k$. Since T is convex, $(\delta_2, \gamma_k^*) \in T$. The slope of segments l_k tends to infinity as k increases. (This follows from the inequality $(\gamma_k - \gamma_1)/(\delta_k - \delta_1) \geq (\gamma_k - \gamma_1)/R$, where R is defined in the formulation of the Lemma.) Then γ_k^* also tends to infinity, and the norm $\|\gamma_k^* Y_0\|$ is unrestricted at the input $\delta_2 X_0$, which contradicts the assumption (A1) about T. ∎

LEMMA 9. *If the PPS T is closed, convex and satisfies assumption* (A1), *then the set Δ^0 defined in* (A.6) *is closed.*

Proof of Lemma 9. Consider any sequence $\{\delta_k\}$, where $\delta_k \in \Delta^0$ for $k = 1, 2, \ldots$, which converges to some δ^*. We need to show that $\delta^* \in \Delta^0$. For every k, there exists a γ_k such that $(\delta_k X_0, \gamma_k Y_0) \in T$. The sequence $\{\delta_k\}$ is a bounded subset of Δ^0. By Lemma 8, the sequence $\{\gamma_k\}$ and, therefore $\{(\delta_k X_0, \gamma_k Y_0)\}$, is also bounded. The latter implies that there exists a subsequence of units $\{(\delta_k X_0, \gamma_k Y_0)\}$ which, from the definition of δ^*, converges to $(\delta^* X_0, \gamma^* Y_0)$ for some γ^*. Since T is a closed set, the limit unit is in T, which implies that $\delta^* \in \Delta^0$. ∎

LEMMA 10. *If the PPS T is closed, convex and satisfies assumption* (A1), *then the function $\bar{\gamma}(\delta)$ is concave and continuous on the set Δ^0 defined in* (A.6).

Proof of Lemma 10. Consider any $\delta_1, \delta_2 \in \Delta^0$ and $\lambda \in [0, 1]$. Then $(\delta_1 X_0, \bar{\gamma}(\delta_1) Y_0) \in T$ and $(\delta_2 X_0, \bar{\gamma}(\delta_2) Y_0) \in T$. (Here definition (8) includes the case $\delta = 0$.) Since T is a convex set, the unit $(\delta^\lambda X_0, \gamma^\lambda Y_0) \in T$ (see definition (A.1). Since $\bar{\gamma}(\delta^\lambda) \geq \gamma^\lambda$, $\bar{\gamma}(\delta)$ is concave on Δ^0.

Further, as a concave function, $\bar{\gamma}(\delta)$ is continuous at any interior point of the interval Δ^0. To prove the continuity at the boundary points, we first prove that the hypograph $hypo\,\bar{\gamma}(\delta) = \{(\delta, \gamma) | \delta \in \Delta^0, \gamma \leq \bar{\gamma}(\delta)\}$ is a closed set. Consider any sequence $\{(\delta_k, \gamma_k)\}$, $k = 1, 2, \ldots$, such that $(\delta_k, \gamma_k) \in hypo\,\bar{\gamma}(\delta)$ for all k, which converges to some pair (δ^*, γ^*).

For every k, $(\delta_k X_0, \bar{\gamma}(\delta_k)Y_0) \in T$ and, by (8), $\gamma_k \leq \bar{\gamma}(\delta_k)$. Since the sequence of δ_k converges, it is bounded and, by Lemma 8, the corresponding sequence of $\bar{\gamma}(\delta_k)$ is also bounded. Then there exists a subsequence of $\{\bar{\gamma}(\delta_k)\}$ which converges to some $\bar{\gamma}^*$. Since T is closed, $(\delta^* X_0, \bar{\gamma}^* Y_0) \in T$. Since $\gamma^* \leq \bar{\gamma}^*$, we have $(\delta^*, \gamma^*) \in$ *hypo* $\bar{\gamma}(\delta)$.

The continuity of function $\bar{\gamma}(\delta)$ on the interval Δ^0 now follows from the fact that $\bar{\gamma}(\delta)$ is concave on Δ^0 and its hypograph is a closed set.[15] ■

LEMMA 11. *Let any function $f(t)$ of a scalar argument t be concave and differentiable on the interval $t \in (a, b)$. Then the function $h(t) = tf'(t) - f(t)$ is not increasing on this interval.*

Proof of Lemma 11. Consider any $t_1, t_2 \in (a, b)$ such that $t_1 < t_2$. We need to prove that $h(t_1) - h(t_2) \geq 0$. Rearranging the terms,

$$h(t_1) - h(t_2) = t_1 \left[f'(t_1) - f'(t_2) \right] + \left[f(t_2) - f(t_1) - f'(t_2)(t_2 - t_1) \right]. \tag{A.7}$$

Since $f(t)$ is concave, the two terms in square parentheses in (A.7) are nonnegative and the lemma follows. (This proof is simpler if the function $f(t)$ is twice differentiable on (a, b), as in this case $h'(t) = f''(t) \leq 0$, and the lemma follows immediately.) ■

Notes

1. In this and similar statements throughout the paper we implicitly assume that the RTS classes can be correctly defined. If the boundary of the technology is given by the production correspondence $F(X, Y) = 0$, the standard conditions needed for the definition of RTS are those of the implicit function theorem. In Section 11 we deal with this issue in greater detail.
2. In the last few years there has been significant interest to the development of non-convex technologies and their practical use. In addition to the well established FDH technology and the non-convex models of Petersen (1990), among the recent developments in this area we find Bogetoft (1996), Bogetoft et al. (2000), Dekker and Post (2001), Kuosmanen (2001) and Podinovski (2005).
   Non-convex technologies are relevant when the divisibility (e.g., time divisibility) of operations cannot be assumed. This may be due to significant start-up times and costs needed for a particular production pattern. Further, even if the divisibility of operations can be accepted, the non-convexity may still be an issue for a variety of different reasons. Agrell et al. (2002) identify a few such reasons, including the specialisation of units and economies of scope, and the different cost of resources depending on quantities. The same view is upheld in the discussion incorporated in Kuosmanen (2003).
3. This equivalence of the local and global classes of RTS is, as highlighted in Note 1, subject to some additional assumptions.
4. In a multiproduct economy each nonzero output vector Y generates the ray γY, where $\gamma > 0$. The ray average cost of vector γY is defined as $c(\gamma Y)/\gamma$ (see, e.g., Baumol et al., 1982). Then Y minimises the ray average cost if the minimum of the ratio $c(\gamma Y)/\gamma$ is attained at $\gamma^* = 1$. The same γ^* maximises the reciprocal value $\gamma / c(\gamma Y)$, which is the average productivity.
5. Podinovski (2004b) highlights an example of a non-convex technology in which the methods of Färe et al. (1983) and Kerstens and Vanden Eeckaut (1999) produce *different* classifications of the types of local RTS, neither of which is correct. The same effect can be observed in our Figure 5.

The method of Färe et al. classes any unit on the boundary strictly between points B and D as exhibiting IRS because their efficiency in the CRS and NIRS reference technologies is the same and, at the same time, smaller than in the original technology T. The method of Kerstens and Vanden Eeckaut classes such units as exhibiting CRS. The true type of local RTS of all such units is DRS. At the same time, their global type is G-SCRS.

6. If the PPS T is a closed set, Theorem 4 becomes a simple corollary to Theorem 5.
7. For our purposes it suffices that DMU_0 is only *R*-efficient in the sense of Definition 5.
8. The value $\delta = 0$ is specifically excluded from Δ because the ray average productivity $\varphi(\delta)$ is undefined at this δ. If the PPS T does not include the zero DMU, the interval Δ is a closed set. Otherwise it is not. For details, see Lemma 9 in the Appendix.
9. This takes into account the fact that the same DMU (X, Y) is represented by different values of δ in the case of a different base unit (13).
10. Obviously, $\varphi^* = 1/m^*$, where m^* is defined by program (4).
11. It is easy to see that $\delta \in \Delta^*$ if and only if unit (10) is a SRU of DMU_0. Since $\bar{\gamma}(\delta) = \varphi^*\delta$ for every $\delta \in \Delta^*$, any SRU of DMU_0 can be represented in the form $(\delta X_0, (\varphi^*\delta)Y_0)$. Therefore, taking into account Lemma 5, the set of all SRUs of DMU_0 is either a linear interval or a single point on the boundary of T.
12. Since Δ^* is a closed subset of the interval Δ, the sets Δ^+ and Δ^- cannot consist of a single point. Therefore, each of these two sets is either empty or has a non-empty interior.
13. For a standard formulation of this theorem see, e.g., Courant (1967). In our context, this theorem requires that (1) the partial derivatives F_δ and $F_{\bar{\gamma}}$ be continuous in some neighbourhood of the point $\delta = \bar{\gamma} = 1$, and (2) $F_{\bar{\gamma}} \neq 0$ at $\delta = \bar{\gamma} = 1$. The proof of Lemma 7 illustrates the use of this theorem.
14. See, e.g., the regularity assumption R3 in Panzar and Willig (1977).
15. For details, see Theorem 7.1 and Corollary 7.5.1 in Rockafellar (1970).

References

Agrell, P. J., P. Bogetoft, M. Brock and J. Tind. (2002). "Efficiency Evaluation with Convex Pairs." KVL, Copenhagen (mimeo).

Ali, A. I. and L. M. Seiford. (1993). "The Mathematical Programming Approach to Efficiency Analysis." In H. O. Fried, C. A. K. Lovell, and S. S. Schmidt (eds.), *The Measurement of Productive Efficiency*. Oxford: Oxford University Press, 120–159.

Banker, R. D. (1984). "Estimating Most Productive Scale Size Using Data Envelopment Analysis." *European Journal of Operational Research* 17, 35–44.

Banker, R. D., H. Chang and W. W. Cooper. (1996). "Equivalence and Implementation of Alternative Methods for Determining Returns to Scale in Data Envelopment Analysis." *European Journal of Operational Research* 89, 473–481.

Banker, R. D., A. Charnes and W. W. Cooper. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30, 1078–1092.

Banker, R. D., W. W. Cooper, L. M. Seiford, R. M. Thrall and J. Zhu. (2004). "Returns to Scale in Different DEA Models." *European Journal of Operational Research* 154, 345–362.

Baumol, W. J., J. C. Panzar and R. D. Willig. (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich, Inc.

Bogetoft, P. (1996). "DEA on Relaxed Convexity Assumptions." *Management Science* 42, 457–465.

Bogetoft, P., J. M. Tama and J. Tind. (2000). "Convex Input and Output Projections of Nonconvex Production Possibility Sets." *Management Science* 46, 858–869.

Briec, W., K. Kerstens, H. Leleu and P. Vanden Eeckaut. (2000). "Returns to Scale on Nonparametric Deterministic Technologies: Simplifying Goodness-of-fit Methods Using Operations on Technologies." *Journal of Productivity Analysis* 14, 267–274.

Charnes, A., W. W. Cooper and E. Rhodes. (1978). "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2, 429–444.

Cooper, W. W., L. M. Seiford and K. Tone. (2000). *Data Envelopment Analysis*. Dordrecht: Kluwer Academic Publishers.

Courant, R. (1967). *Differential and Integral Calculus*. Volume II, London: Blackie and Son Ltd.

Dekker, D. and T. Post. (2001). "A Quasi-concave DEA Model with an Application for Branch Performance Evaluation." *European Journal of Operational Research* 132, 296–311.

Deprins, D., L. Simar and H. Tulkens. (1984). "Measuring Labor-efficiency in Post Offices." In M. Marchand, P. Pestieau, and H. Tulkens (eds.), *The Performance of Public Enterprises*. Amsterdam: Elsevier Science Publishers, pp. 243–267.

Färe, R., S. Grosskopf and C. A. K. Lovell. (1983). "The Structure of Technical Efficiency." *Scandinavian Journal of Economics* 85, 181–190.

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

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

Kerstens, K. and P. Vanden Eeckaut. (1999). "Estimating Returns to Scale Using Non-parametric Deterministic Technologies: A New Method Based on Goodness-of-fit." *European Journal of Operational Research* 113, 206–214.

Kuosmanen, T. (2001). "DEA with Efficiency Classification Preserving Conditional Convexity." *European Journal of Operational Research* 132, 326–342.

Kuosmanen, T. (2003). "Duality Theory and Non-convex Technologies ." *Journal of Productivity Analysis* 20, 273–304.

Panzar, J. C. and R. D. Willig. (1977). "Economies of Scale in Multi-output Production." *Quarterly Journal of Economics* 91, 481–491.

Petersen, N. C. (1990). "Data Envelopment Analysis on a Relaxed Set of Assumptions." *Management Science* 36, 305–314.

Podinovski, V. V. (2002). "Global Scale Characteristics in DEA." Warwick Business School Research Paper No. 353, University of Warwick, UK.

Podinovski, V. V. (2004a). "On the Linearisation of Reference Technologies for Testing Returns to Scale in FDH Models." *European Journal of Operational Research* 152, 800–802.

Podinovski, V. V. (2004b). "Local and Global Returns to Scale in Performance Measurement." *Journal of the Operational Research Society* 55, 170–178.

Podinovski, V. V. (2005). "Selective Convexity in DEA Models." Forthcoming in the *European Journal of Operational Research*.

Rockafellar, R. T. (1970) *Convex Analysis*. Princeton, NJ: Princeton University Press.

Seiford, L. M. and J. Zhu. (1999). "An Investigation of Returns to Scale in Data Envelopment Analysis." *Omega* 27, 1–11.

Thanassoulis, E. (2001). *Introduction to the Theory and Application of Data Envelopment Analysis*. Dordrecht: Kluwer Academic Publishers.