European Journal of Operational Research 259 (2017) 576–586

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Global and local scale characteristics in convex and nonconvex nonparametric technologies: A first empirical exploration☆

Giovanni Cesaroni[a], Kristiaan Kerstens[b,*], Ignace Van de Woestyne[c]

[a] *Prime Minister's Office, Department for economic policy, Via della Mercede 9, Rome I-00187, Italy*
[b] *CNRS-LEM (UMR 9221), IESEG School of Management, 3 rue de la Digue, Lille F-59000, France*
[c] *Research Unit MEES, KU Leuven, Warmoesberg 26, Brussel B-1000, Belgium*

ARTICLE INFO

Article history:
Received 20 January 2015
Accepted 18 October 2016
Available online 21 October 2016

Keywords:
Data Envelopment Analysis
Free Disposal Hull
Global returns to scale
Local returns to scale

ABSTRACT

The purpose of this contribution is to empirically implement and supplement the proposals made by Podinovski (2004b) to explore the nature of both global and local returns to scale in nonconvex nonparametric technologies. In particular, we both propose a simplified method to compute the global returns to scale and employ some secondary data sets to investigate the frequency of the special case of global sub-constant returns to scale. Furthermore, when determining global returns to scale using both convex and nonconvex technologies, we verify how often the resulting information is concordant or conflicting. Finally, besides comparing the FDH and DEA evolution of ray-average productivity for some typical individual observations, we introduce in the literature two original methods for the determination of local returns to scale in nonconvex technologies.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Thanks to the seminal article of Charnes, Cooper, and Rhodes (1978), the nonparametric approach to production theory has become one of the success stories in the operations research (OR) literature in terms of both methodological developments and empirical applications. While one of the early bibliographical overview article listed about 800 published articles and dissertations related to Data Envelopment Analysis (DEA) over the years 1978–1996 (see Seiford, 1997), one of the more recent bibliography articles of Emrouznejad, Parker, and Tavares (2008) counted already 4000 research articles in journals or book chapters up to the year 2007.[1]

While the axiom of convexity is traditionally maintained in these nonparametric production models (see Afriat, 1972; Banker, Charnes, & Cooper, 1984; Charnes et al., 1978; Diewert & Parkan, 1983 or any of the early contributions in both economics and OR), Afriat (1972) was probably the first to mention a basic single output nonconvex technology imposing the assumptions of free disposal of inputs and outputs. Its multiple output extension has probably first been proposed in Deprins, Simar, and Tulkens (1984) and these authors introduced the moniker Free Disposal Hull (FDH).[2]

Convexity is justified for time divisible technologies (see Hackman, 2008), but becomes questionable when time indivisibilities compound all other reasons for spatial nonconvexities (e.g., indivisibilities, increasing returns to scale, economies of specialization, externalities, etc.). Shephard (1967, p. 215) puts things clearly when discussing the axiom of quasi-concavity of the production function in relation to convexity of the input level sets when formally defining the notion of a production function:

> The last one is effectively the only assumption which would appear to be restrictive, but even so it is essential if the production function is to represent the maximum output obtainable for time divisible processes. If the processes are not time divisible, the input $[(1-\theta)x+\theta y]$ is not evidently feasible. .... We exclude considerations of such technologies.

In addition to this general criticism, there are other more specific criticisms of convexity around in the literature. For instance, Emrouznejad and Amin (2009) indicate that the traditional

☆ We acknowledge comments made by participants in the North American Productivity Workshop VIII (Ottawa) and in the 10th Asia-Pacific Productivity Conference (Brisbane). The constructive comments of three referees are gratefully acknowledged.

* Corresponding author.
E-mail address: k.kerstens@ieseg.fr (K. Kerstens).

[1] Including unpublished dissertations, working papers, and conference papers would have led to over 7000 entries.

[2] Tone and Sahoo (2003, p. 172) mention Scarf (1981a, 1981b) as an important but neglected predecessor of FDH, because he studied activity analysis models based on integer data.

http://dx.doi.org/10.1016/j.ejor.2016.10.030
0377-2217/© 2016 Elsevier B.V. All rights reserved.

convexity axiom is problematic when some of the inputs and/or some of the outputs are ratio variables.

This basic FDH model has been extended in at least two directions. First, Kerstens and Vanden Eeckaut (1999) introduced constant, nonincreasing and nondecreasing returns to scale technologies complementary to the assumption of flexible or variable returns to scale embodied in the basic FDH model. Furthermore, these same authors proposed a new goodness-of-fit method to infer the characterization of global returns to scale for nonconvex technologies, since none of the existing methods (see, e.g., Seiford & Zhu, 1999 for an early overview and Banker, Cooper, Seiford, Thrall, & Zhu, 2004 for a more recent version) was suitable in this nonconvex setting. Second, this family of nonconvex technologies has been supplemented by nonconvex cost functions with corresponding returns to scale assumptions in Briec, Kerstens, and Vanden Eeckaut (2004).[3]

While these nonconvex technology and cost models are nowhere as popular as the convex DEA counterparts, the basic FDH model and its extensions have been regularly applied to assess performance-related research questions in a variety of sectors. We offer a limited selection of examples to provide some flavor of these results. Alam and Sickles (1998) study the evolution of technical efficiency in the US airline industry and analyze the news value of changes in frontier performance in relation to the stock market prices. Destefanis (2003) analyzes the macroeconomic relationship between the growth of output and the growth of productivity (known as Verdoorn's law) using nonconvex FDH models. Tone and Sahoo (2003) argue and illustrate that the nonconvex FDH model applied to a multi-stage production technology is capable to capture scale effects arising from process indivisibilities, whereas standard convex nonparametric technologies fail to exhibit such scale effects. Cummins and Zi (1998) contrast convex and nonconvex estimates of both technical and cost efficiency for US life insurers, while Balaguer-Coll, Prior, and Tortosa-Ausina (2007) document cost efficiency differences among Spanish municipalities.

An important point to note is that the results of these nonconvex technology and cost frontiers often yield different results compared to the convex ones. While it is true that nonconvex technology frontiers lead to higher efficiency levels and more efficient observations, the studies of Balaguer-Coll et al. (2007) and Cummins and Zi (1998) document convincingly that convex cost frontier estimates may be substantially below the nonconvex ones under variable returns to scale.

Podinovski (2004a, 2004b) is the first to indicate that the goodness-of-fit method of Kerstens and Vanden Eeckaut (1999) to characterize global returns to scale for nonconvex technologies – which just like Färe, Grosskopf, and Lovell (1983) uses only scale efficiency measures – is incomplete. In particular, he argues that one must distinguish a fourth type of global sub-constant returns to scale case in addition to the three traditional cases (constant, decreasing and increasing returns to scale). This global sub-constant returns to scale case allows a unit to achieve its most productive scale size (see Banker et al., 1984) by both reducing and increasing its scale of operations. This fourth type of global sub-constant returns to scale can never occur in traditional convex DEA technologies.

Independent of this contribution, there have been three articles that basically simplify the computations needed to implement the goodness-of-fit method of Kerstens and Vanden Eeckaut (1999) to characterize returns to scale: Soleimani-damaneh, Jahanshahloo, and Reshadi (2006), Soleimani-damaneh and Reshadi (2007), and Soleimani-damaneh and Mostafaee (2009). In fact, Soleimani-damaneh and Mostafaee (2009) furthermore offer some stability intervals to preserve the returns to scale classification via a polynomial time algorithm based on combining certain ratios of inputs and outputs. However, the classification procedure for global returns to scale proposed by these authors does not allow for the sub-constant returns to scale case. Therefore, we discuss how to amend their procedures for this purpose.

As far as the role of local returns to scale is concerned, Banker (1984) and especially Banker and Thrall (1992) show that in a convex technology global and local characterizations – based on scale efficiency and scale elasticity measures, respectively – coincide.[4] The innovation of Podinovski (2004a, 2004b) is that he points out that this equivalence between global and local indicators breaks down for nonconvex technologies, due to the non-monotonic behavior of the ray average productivity (RAP) of a unit when expanding or contracting towards a point of most productive scale size.[5] However, he only provides an illustration of the RAP function in a single input and output fictitious FDH technology (see Podinovski, 2004a, p. 234), while our aim is to depict this behavior and its consequences in an empirical multiple inputs and outputs setting. Moreover, we try to fill a gap in the literature due to the lack of explicit methods for ascertaining local returns to scale in nonconvex technologies, with the aim of complementing the work of Podinovski (2004a, 2004b) even from a theoretical point of view.

This contribution intends to achieve several goals. First, we want to empirically determine the prevalence of the global sub-constant returns to scale case. Second, we want to establish some specific links between the Podinovski (2004a, 2004b) articles on the one hand, and the contributions made by Soleimani-damaneh et al. (2006) and Soleimani-damaneh and Reshadi (2007) on the other hand. Third, we want to explore the similarities and differences between global returns to scale characterizations under the hypothesis of convexity or nonconvexity. Fourth, we shed some light on the changes in returns to scale in an empirical multiple inputs and outputs nonconvex technology by depicting the evolution of ray-average productivities for a selection of particular observations, and by comparing this evolution to its convex counterpart. Finally, we propose two methods for the classification of local returns to scale in nonconvex technologies and discuss their basic properties and relationship to the traditional criterion based on scale elasticity. To the best of our knowledge, this is the first contribution shedding some light on these issues.

For these purposes, this paper is structured as follows. Section 2 provides some basic definitions of the traditional convex and the less widely applied nonconvex technologies. Section 3 summarizes the known results to characterize returns to scale at the global level and introduces two criteria for the determination of their local counterpart in FDH and in a nonconvex smooth technology. Then follows a Section 4 with some empirical illustrations based on secondary data sets. Section 5 concludes and outlines future research issues.

2. Nonparametric technologies: a unified representation

Consider a set of K observations $A = \{(x_1, y_1), \ldots, (x_K, y_K)\} \in \mathbb{R}_+^{m+n}$. A production technology describes all available possibilities to transform input vectors $x = (x_1, \ldots, x_m) \in \mathbb{R}_+^m$ into output vectors $y = (y_1, \ldots, y_n) \in \mathbb{R}_+^n$. The production possibility set or

[3] Ray (2004) shows that the nonconvex cost function based on flexible returns to scale FDH is the multiple output version of the cost function implicit in the Weak Axiom of Cost Minimization of Varian (1984).

[4] As shown in Appendix A, the equivalence holds for standard (i.e., one-stage) convex production technologies. It may not hold for more complex technologies, such as the two-stage examples discussed in Sahoo, Zhu, Tone, and Klemen (2014). In this article, we restrict attention to standard, one-stage production technologies.

[5] RAP indicates average productivity in a multiple inputs and output technology. See also Ray (2004, pp. 63–64) for this RAP notion.

technology S summarizes the set of all feasible input and output vectors: $S = \{(x, y) \in \mathbb{R}_+^{m+n} : x \text{ can produce } y\}$. Given our focus on input-oriented efficiency measurement later on, this technology can be represented by the input correspondence $L : \mathbb{R}_+^n \to 2^{\mathbb{R}_+^m}$ where $L(y)$ is the set of all input vectors that yield at least the output vector y:

$$L(y) = \{x : (x, y) \in S\}. \tag{1}$$

The radial input efficiency measure can be defined as:

$$E_i(x, y) = \min\{\lambda \ : \ \lambda \geq 0, \ \lambda x \in L(y)\}. \tag{2}$$

This Farrell efficiency measure, which is the inverse of the input distance function, indicates the minimum contraction of an input vector by a scalar λ while still remaining in the input correspondence. Obviously, the resulting input combination is located at the boundary of this input correspondence. For our purpose, the radial input efficiency has two key properties (see, e.g., Hackman, 2008). First, it is smaller or equal to unity ($0 < E_i(x, y) \leq 1$), whereby efficient production on the isoquant of $L(y)$ is represented by unity and $1 - E_i(x, y)$ indicates the amount of inefficiency. Second, it has a cost interpretation.

Nonparametric specifications of technology can be estimated by enveloping these K observations in the set A while maintaining some basic production axioms (see Hackman, 2008 or Ray, 2004). We are interested in defining minimum extrapolation technologies satisfying strong disposability in the inputs and outputs, all four traditional returns to scale hypotheses (i.e., constant, nonincreasing, nondecreasing and variable (flexible) returns to scale), including those technologies that satisfy the assumption of convexity and those that do not

A unified algebraic representation of convex and nonconvex technologies under different returns to scale assumptions for a sample of K observations is found in Briec et al. (2004):

$$S^{\Lambda,\Gamma} = \left\{ (x, y) \in \mathbb{R}_+^{m+n} : \ x \geq \sum_{k=1}^{K} x_k \alpha \, z_k, \ y \leq \sum_{k=1}^{K} y_k \alpha \, z_k, \right.$$
$$\left. \sum_{k=1}^{K} z_k = 1, \ z_k \in \Lambda, \alpha \in \Gamma \right\}, \tag{3}$$

where

(i) $\Gamma \equiv \Gamma^{\text{CRS}} = \{\alpha : \ \alpha \geq 0\}$;
(ii) $\Gamma \equiv \Gamma^{\text{NDRS}} = \{\alpha : \ \alpha \geq 1\}$;
(iii) $\Gamma \equiv \Gamma^{\text{NIRS}} = \{\alpha : \ 0 \leq \alpha \leq 1\}$;
(iv) $\Gamma \equiv \Gamma^{\text{VRS}} = \{\alpha : \ \alpha = 1\}$; and
(i) $\Lambda \equiv \Lambda^{\text{C}} = \{z_k \geq 0\}$, and (ii) $\Lambda \equiv \Lambda^{\text{NC}} = \{z_k \in \{0, 1\}\}$.

First, there is the activity vector (z) operating subject to a convexity (C) or nonconvexity (NC) constraint. Second, there is a scaling parameter (α) allowing for a particular scaling of all K observations spanning the technology. This scaling parameter is smaller than or equal to 1 or larger than or equal to 1 under nonincreasing returns to scale (NIRS) and nondecreasing returns to scale (NDRS) respectively, fixed at unity under variable returns to scale (VRS), and free under constant returns to scale (CRS).

Briefly discussing the computational methods for obtaining the radial input efficiency measure (2) for each evaluated observation relative to all technologies in (3), the convex case just requires solving a nonlinear programing problem (NLP): this is evidently simplified to the familiar linear programing (LP) problem found in the literature (see Hackman, 2008 or Ray, 2004) by substituting $w_k = \delta z_k$. For nonconvex technologies, nonlinear mixed integer programs must be solved in (3): however, Podinovski (2004c), Leleu (2006) and Briec et al. (2004) propose mixed integer programs, LP problems, and closed form solutions derived from an implicit enumeration strategy, respectively. Kerstens and Van de Woestyne (2014) review all methods in this nonconvex case in more detail and empirically document that implicit enumeration is by far the fastest solution strategy.

3. Characterizing returns to scale

3.1. Global returns to scale

For a given input mix and given output mix a Most Productive Scale Size (MPSS) point refers to a scale size where the level of outputs produced 'per unit' of the inputs is maximized. Following Banker (1984), Banker et al. (1984, p. 37) and Banker and Thrall (1992, Definition 1), the MPSS notion can be defined as follows.

Definition 3.1. A production possibility $(x_M, \ y_M) \in S^{\Lambda, \ VRS}$ represents an MPSS point if and only if for all production possibilities $(\delta x_M, \ \gamma y_M) \in S^{\Lambda, \ VRS}$ we have $\gamma / \delta \leq 1$.

This notion of MPSS is key in determining returns to scale for general technologies, since it does not require any differentiability assumption (in contrast to the scale elasticity notion). Note that Podinovski (2004a, Definition 2) defines MPSS as the inverse of the above ratio.

As a direct consequence of this definition, $(x_M, \ y_M) \in S^{\Lambda, \ VRS}$ represents an MPSS point if and only if $r^* = 1$ with

$$r^* = \max\left(\frac{\gamma}{\delta} : (\delta x_M, \gamma y_M) \in S^{\Lambda, VRS}, \ \delta, \gamma > 0\right). \tag{4}$$

This implies that at the optimum, $r^* = 1 \Leftrightarrow \gamma^* = \delta^*$, which reflects the familiar condition for proportional changes in inputs to equal proportional changes in outputs at the optimum.

Banker (1984) shows that in a convex technology each scale-efficient point (i.e., CRS efficient) is an MPSS and also the reverse (see Banker, 1984, Proposition 2), while each scale-inefficient point locally exhibits either decreasing or increasing returns to scale according to the sign of the divergence between their actual scale size and their MPSS (see Banker, 1984, Corollary 1). Thus, a classification method can exclusively rely on the "global" comparison between a unit and its MPSS (i.e., its scale efficiency), without depending explicitly on the quantitative information supplied by the "local" scale elasticity measure.

In the literature, several methods are available to obtain qualitative information regarding global returns to scale (see Seiford & Zhu, 1999). Since none of these existing methods are suitable for nonconvex technologies, Kerstens and Vanden Eeckaut (1999, Proposition 2) generalize the existing goodness-of-fit method proposed by Färe et al. (1983) in a convex setting such that it becomes perfectly general. Obviously, this qualitative information holds for efficient points only: these are either efficient observations, or projection points in case of initially inefficient observations.

Proposition 3.1. *Using $E_i(x, y|.)$ and conditional on an efficient point, technology $S^{\Lambda, \ VRS}$ is characterized by:*

(a) *GCRS* $\Leftrightarrow E_i(x, y|CRS) = \max\{E_i(x, y|CRS), E_i(x, y|NIRS), E_i(x, y|NDRS)\}$.
(b) *GIRS* $\Leftrightarrow E_i(x, \ y|NDRS) > \max\{E_i(x, \ y|CRS), \ E_i(x, \ y|NIRS)\}$.
(c) *GDRS* $\Leftrightarrow E_i(x, \ y|NIRS) > \max\{E_i(x, \ y|CRS), \ E_i(x, \ y|NDRS)\}$.

where GCRS, GIRS and GDRS stand for globally constant, increasing and decreasing returns to scale, respectively.

As noted by Podinovski (2004b, p. 173), following Briec, Kerstens, Leleu, and Vanden Eeckaut (2000, Proposition 5) one can simplify the above result for general (i.e., convex and nonconvex) technologies.

Proposition 3.2. *Using $E_i(x, y|.)$ and conditional on an efficient point, technology $S^{\Lambda,\ VRS}$ is characterized by:*

(a) $GCRS \Leftrightarrow E_i(x, y|NIRS) = E_i(x, y|NDRS)$.
(b) $GIRS \Leftrightarrow E_i(x, y|NDRS) > E_i(x, y|NIRS)$.
(c) $GDRS \Leftrightarrow E_i(x, y|NIRS) > E_i(x, y|NDRS)$.[6]

This result is qualified by Podinovski (2004a, Theorem 3) and Podinovski (2004b, Theorem 2) in that he adds a fourth case of global sub-constant returns to scale case that is only relevant for nonconvex technologies.

Proposition 3.3. *Using $E_i(x, y|.)$ and conditional on an efficient point, technology $S^{\Lambda,\ VRS}$ is characterized by:*

(a) $GCRS \Leftrightarrow E_i(x, y|NIRS) = E_i(x, y|NDRS) = E_i(x, y|VRS)$.
(b) $GIRS \Leftrightarrow E_i(x, y|NIRS) < E_i(x, y|NDRS) \le E_i(x, y|VRS)$.
(c) $GDRS \Leftrightarrow E_i(x, y|NDRS) < E_i(x, y|NIRS) \le E_i(x, y|VRS)$.
(d) $GSCRS \Leftrightarrow E_i(x, y|NIRS) = E_i(x, y|NDRS) < E_i(x, y|VRS)$.

where GSCRS stands for the global sub-constant returns to scale case.

As stressed in Podinovski (2004a, 2004b), this case of global sub-constant returns to scale cannot occur in convex technologies. Instead of solving for these three efficiency measures using any of the solution methods listed above, we follow a specific theorem in Soleimani-damaneh et al. (2006, p. 1057) that proposes a simple enumeration algorithm valid for nonconvex technologies solely to guarantee a maximal computational advantage.

Proposition 3.4. *For a given FDH-efficient observation (x_o, y_o), i.e. $E_i(x_o, y_o|VRS) = 1$, let $\lambda^{jo} = \max\{\frac{y_{ro}}{y_{rj}} : 1 \le r \le n, y_{ro} + y_{rj} > 0\}$ and $\theta^{jo} = \max\{\frac{x_{ij}\lambda^{jo}}{x_{io}} : 1 \le i \le m, x_{io} + x_{ij} > 0\}$ for $j = 1, \ldots, K$. Let $E_i^{NC}(x_o, y_o|CRS) = \min\{\theta^{jo} : j = 1, \ldots K\}$. Now denote the set $A_o = \{k \in \{1, \ldots, K\} : \theta^{ko} = E_i^{NC}(x_o, y_o|CRS)\}$. Assuming that (x_o, y_o) is an FDH-efficient point, then the following conditions identify the situation of RTS at this point:*

(a) *There exists $k \in A_o$ such that $\lambda^{ko} = 1 \Rightarrow GCRS$.*
(b) *$\lambda^{ko} < 1$ for each $k \in A_o \Rightarrow GIRS$.*
(c) *$\lambda^{ko} > 1$ for each $k \in A_o \Rightarrow GDRS$.*
(d) *$\lambda^{ko} \neq 1$ for each $k \in A_o$ and furthermore, there exist $k, k' \in A_o$ such that $\lambda^{ko} < 1$ and $\lambda^{k'o} > 1 \Rightarrow GSCRS$.*

Proof: For cases (a)–(c) see Soleimani-damaneh et al. (2006, p. 1058). The intuition for (d) can be given as follows. Expression (10) in Cesaroni and Giovannola (2015, p. 124) shows that θ^{ko} is determined by the global maximum of the RAP of the FDH-efficient observation (x_o, y_o). Then, it can be understood that case (d) above occurs when RAP is maximized at both the left ($\lambda^{ko} < 1$) and the right ($\lambda^{ko} > 1$), but not at the efficient point itself. □

Since Soleimani-damaneh et al. (2006) have not considered the possibility of global sub-constant returns to scale, which corresponds to their case (d), we have extended this proposition and labeled the outcome with *GSCRS*, because of the presence of scale inefficiency in the DMU under evaluation (inefficiency which the authors fail to consider). Exactly the same improvement applies to Soleimani-damaneh and Reshadi (2007, Theorem 1) and Soleimani-damaneh and Mostafaee (2009, Theorem 1).

To the best of our knowledge, no article ever reported any empirical evidence on the incidence of the global sub-constant returns to scale in relation to the other cases.

[6] This proposition qualifies Briec et al. (2000, Proposition 4): as an implication of their Proposition 5, since a CRS technology is always the union of NIRS and NDRS hulls, the goodness-of-fit test in their Proposition 2 always simplifies (not just for convex technologies).

3.2. Local returns to scale

The exact relation between scale efficiency, which involves the global maximization of RAP, and scale elasticity, which is based on the maximization of RAP in a small neighborhood, has first been elaborated in convex nonparametric production frontiers in the seminal analysis of Banker and Thrall (1992). These authors prove explicitly the equivalence between the local method based on the values of scale elasticity and the global method relying on the sign of the difference between actual and most productive scale sizes (see Banker and Thrall, 1992, Propositions 3 and 4, resp.). Other contributions on this topic are, among others, those of Førsund and Hjalmarsson (2004) and Førsund, Hjalmarsson, Krivonozhko, and Utkin (2007). The potential empirical differences between both these concepts have been illustrated in, for instance, Evanoff and Israilevich (1995).

However, as noted by Podinovski (2004a, p. 228): “in a general nonconvex technology the RTS classes no longer play the role of global indicators”, because local maxima of the RAP function are neither necessarily global maxima nor necessarily located in the same direction. In other words, even for a differentiable nonconvex technology, global analysis of returns to scale must be separated from local analysis (i.e. “RTS classes”). With regard to the latter, Podinovski (2004b, p. 172 and p. 177) clearly points out that the use of the traditional notion of scale elasticity is only possible for technologies with a sufficiently smooth boundary, but that this notion is undefined for FDH because of the discontinuity of the average productivity function at any efficient point (x_0, y_0).

In fact, the DEA frontier is sufficiently smooth to ensure the continuity of the average productivity function at (x_0, y_0), which permits the use of the criterion based on the interval determined by right-hand and left-hand scale elasticities, $SE^+ = \lim_{\delta \to 1^+} \frac{\gamma(\delta) - 1}{\delta - 1}$ and $SE^- = \lim_{\delta \to 1^-} \frac{\gamma(\delta) - 1}{\delta - 1}$ with $SE^+ \le SE^-$ (see Banker, 1984; Hadjicostas & Soteriou, 2006). According to this criterion, we have local CRS if 1 belong to this interval, local IRS if $SE^+ > 1$ and local DRS if $SE^- < 1$. The three cases describe a situation where RAP in a marginally small neighborhood of (x_0, y_0) is: maximized at the efficient point, maximized at the right and left end of the neighborhood, respectively.

However, the scale elasticity approach is not suitable to an FDH technology. In fact, here, at any (x_0, y_0) we have $SE^+ = 0$ and $SE^- = \infty$. But, the drop (discontinuity) in RAP prevents from classifying this case as local DRS because RAP is not necessarily maximized at the left end of the neighborhood. This argument is clearly illustrated in Fig. 1, where RAP r^* and δ are displayed on the vertical and horizontal axis respectively, and A is a non-GCRS efficient point under examination having coordinates (1, 1).

Two important characteristics shown in this Fig. 1 must be pointed out. First, in the open interval $(1-\varepsilon, 1+\varepsilon)$, or in any smaller interval, RAP is maximized at the efficient point A. Second, the half-closed interval $(1-\varepsilon, 1+\varepsilon]$, having the same size as the former interval, contains the efficient point B which determines a RAP greater than 1 (while not necessarily being the MPSS of A). With regard to this, it must also be considered that in multiple-input and multiple-output applications efficient points like A, B and C can locate very close to each other, so that ε may turn out to be negligible (see Section 4.4).

The first consideration shows that, in an FDH technology, it is in principle possible to classify each efficient point as exhibiting local CRS, within an interval whose size varies across classified points. However, this is a sheer consequence of the discontinuity in the RAP function. In fact, the second consideration clarifies that there may exist intervals in which a marginally small resizing results in an immediate improvement of RAP, thereby invalidating

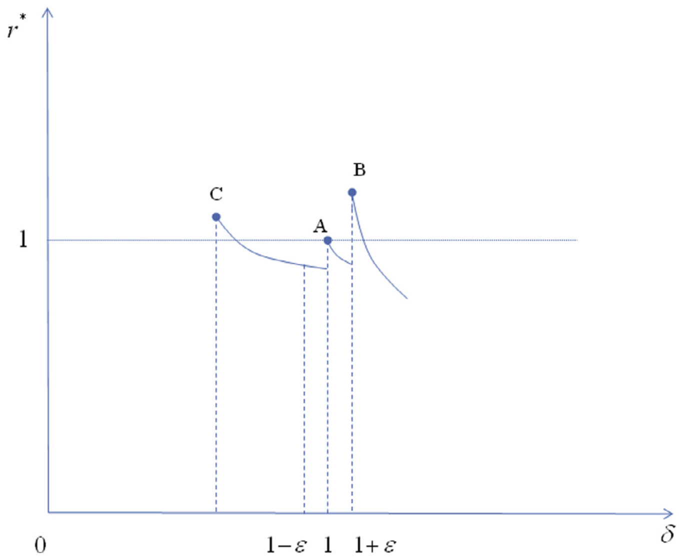

Fig. 1. Relation between RAP (r^*) and δ.

the local CRS characterization. Clearly, these problems do not arise with an GCRS efficient-point, because here RAP reaches its global maximum (if A were such a point, in Fig. 1 B and C would be on or below the straight line passing through 1): GCRS efficient points can be correctly classified as exhibiting local CRS.

The following method is then proposed for the local classification of an efficient point (x_0, y_0) in an FDH technology:

Definition 3.2. Define as relative maxima of the RAP of (x_0, y_0) the frontier points which yield RAP $>$ 1, and choose an exogenous small size $\varepsilon > 0$ for the neighborhood of the efficient point, then in the closed interval $[\,1-\varepsilon, 1+\varepsilon\,]$ we have:

1. Local CRS if there is no relative maximum of RAP.
2. Local IRS if the greatest relative maximum of RAP is in $(1, 1+\varepsilon\,]$.
3. Local DRS if the greatest relative maximum of RAP is in $[\,1-\varepsilon, 1)$.
4. Local SCRS if the greatest relative maximum of RAP in $[\,1-\varepsilon, 1)$ is equal to that in $(1, 1+\varepsilon\,]$.

Following this proposed definition, Fig. 1 illustrates the case of local IRS.

Observe that Definition 3.2 extends Podinovski's global taxonomy to the local level and that, for a given point, these two classifications can in practice diverge. As for the latter feature, note in fact that in a nonconvex technology a local maximum (i.e., the greatest relative maximum of RAP in the ε neighborhood) does not necessarily coincide with the global maximum (MPSS) and, moreover, it is not necessarily located in the same direction.

Furthermore, two remarks are in order. First, the method is a refinement of the uniform local-CRS classification otherwise delivered by the discontinuity of RAP in the FDH technology: the greater ε the more it is likely that a number of DMUs abandon the CRS characterization. Second, the suggested method has the advantage of detecting the point with highest RAP in an interval which managers and regulators rate as feasible given financial constraints and adjustment costs that may limit the size of short-run adjustments in the scale of operations. This feature would not necessarily be maintained in a classification criterion based on the endogenous choice of the neighborhood, such as it is the case of a symmetric interval determined by the efficient-point which is nearest to (x_0, y_0) (see Fig. 7 in Section 4.4).

Having presented some basic properties and interpretation of the local classification criterion proposed in Definition 3.2, we finally address two important issues regarding the nature of the method and its extension to technologies different from FDH.

We point out that the principle on which the method proposed in Definition 3.2 is based is the same underlying the standard notion of local RTS, i.e., the maximization of RAP in a small neighborhood of an efficient point under examination. As a consequence, it is intuitive that our local criterion could apply to different technologies, such as DEA and smooth nonconvex technologies where continuity of the RAP function allows to abandon the neighborhood of exogenous small size. In DEA, the application of Definition 3.2 yields the method based on right- and left-scale elasticities. In fact, to ascertain this, it suffices to observe that in a convex technology case 4 is impossible, while cases 1–3 can exactly be associated to the specific reference-values of the scale elasticities discussed above at p. 4. Under nonconvexity, we propose Definition 3.3 to extend our local criterion to a smooth analogue of the FDH technology.

Definition 3.3. In a smooth nonconvex technology the local RTS at an efficient point (x_0, y_0) are characterized as follows:

1. Local CRS if $SE = 1$ and the second order derivative of RAP is negative.
2. Local IRS if $SE > 1$.
3. Local DRS if $SE < 1$.
4. Local SCRS if $SE = 1$ and the second order derivative of RAP is positive.

It can be easily seen that each of the four cases envisaged by Definition 3.3 is the straightforward application of the corresponding case of Definition 3.2 to a marginally small neighborhood of (x_0, y_0). For an illustration of the application of this criterion, one can consult Fig. 5 in Podinovski (2004a, p. 235). In this figure, point K is local CRS, point A is local IRS, point F is local DRS, and point D is local SCRS.

To the best of our knowledge, Definitions 3.2 and 3.3 introduce in the literature two original methods for the local classification of RTS in general nonconvex technologies. Albeit new, these methods are indeed based on the same standard principle of local RTS, i.e. the maximization of RAP in a small neighborhood of an efficient point.

3.3. Production frontiers

The reconstruction of production frontiers has been analyzed in a few contributions (see, e.g., Hackman, 2008, Chapter 10 for a brief review). Since the convex technologies in (3) are convex polyhedra, facets can be enumerated so as to reconstruct the boundaries of the technology. A two-dimensional projection is then defined relative to a particular point of the technology. For example, Krivonozhko, Utkin, Volodin, Sablin, and Patrin (2004) offers parametric optimization tools to reconstruct an intersection of the multidimensional convex production frontier with a two-dimensional plane determined by any pair of given directions. We simply adapt the same idea to a nonconvex technology. Moreover, to the best of our knowledge, this is the first time that a computed section of an FDH technology is ever displayed.

While for the nonconvex case we follow the same basic setup, we employ a specific enumeration algorithm. Indeed, as Podinovski (2004a, p. 233) indicates, MPSS points can be determined by solving either for the MPSS definition (3.1) relative to a VRS technology ($S^{\Lambda,\,VRS}$), or a radial efficiency measure relative to a CRS technology ($S^{\Lambda,\,CRS}$) (see also Banker, 1984, Proposition 1). Following Soleimani-damaneh and Reshadi (2007, Lemma 1), the former solution is equivalent to the specific enumeration algorithm

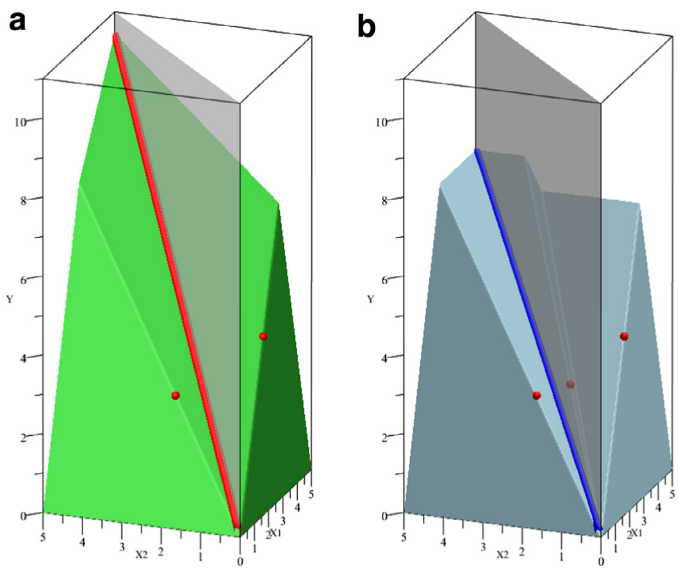

Fig. 2. Two inputs single output (a) convex and (b) nonconvex CRS technology.

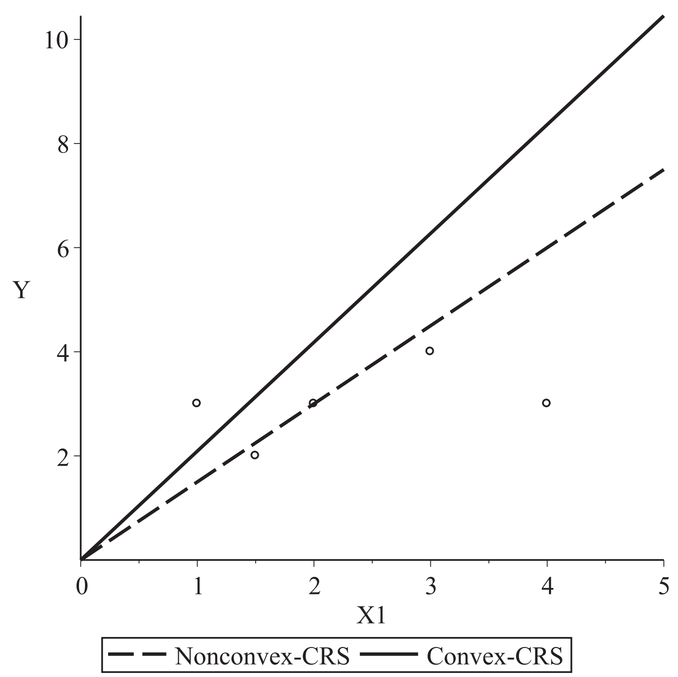

Fig. 3. Single input single output representation of section of convex and nonconvex CRS technologies.

developed in Soleimani-damaneh et al. (2006, p. 1057) and Soleimani-damaneh and Reshadi (2007, pp. 2172–2173) for nonconvex technologies and it is based on the notion of RAP.

It is important to point out that average productivity under convexity may be higher or equal to average productivity under nonconvexity. To develop this intuition, one can look at the two Figs. 2 and 3.

From a small numerical example we reconstructed in Fig. 2 both a convex (part (a)) and nonconvex (part (b)) CRS technology in a two inputs single output space. In the convex case, only two points span the three faces of the convex cone. In the nonconvex case, three observations span the ridge lines emanating from the origin and determining the nonconvex cone because these observations operate under CRS. Based on these 3-D figures one may infer that the convex cone contains the nonconvex cone.

This is clearly made visible by the section with a vertical plane along a ray through the origin and along the single output depicted in the same Fig. 2. Fig. 3 depicts this latter section in just two dimensions by a projection into the X_1Y-plane: it is clear that average productivity under convexity is higher than under nonconvexity along this particular section of Fig. 2.

4. Empirical illustrations

In this section, we first present the data sets adopted from existing studies. Then, we present empirical results on global returns to scale. Thereafter, we turn to a selection of results focusing on local returns to scale.

4.1. Secondary data sets employed

To empirically illustrate these developments, we employ several existing data sets. Table 1 summarizes some key features of each data set: sample size, number of inputs and outputs, and the sector. There is one small unbalanced panel (Färe, Grosskopf, & Logan, 1983) and four cross sections (Cesaroni, 2011; Fan, Li, & Weersink, 1996; Färe, Grosskopf, Logan, & Lovell, 1985; Haag, Jaska, & Semple, 1992). Note that the time dimension in the panel is ignored: this amounts to assuming there is no technical change over the five time periods.

The main points to note are the following. There are three single output samples, and two multiple-output samples. Sample sizes vary from very small to rather big. The data sets have been sorted in Table 1 according to their sample size. In the other tables we maintain the same order.

4.2. Global returns to scale

Turning to the determination of global returns to scale, we set ourselves two goals. First, we want to document any eventual differences between convex and nonconvex technologies in terms of the nature of returns to scale for individual observations. This has to the best of our knowledge nowhere been reported. Second, it is important to evaluate the incidence of the global sub-constant returns to scale case developed by Podinovski (2004a, 2004b).

Table 2 reports the basic decomposition of overall technical efficiency (OTE) into a scale efficiency (SCE) and a technical efficiency (TE) component. This amounts to comparing efficiency relative to CRS and VRS technologies. In particular, $OTE = E_i(x, y|CRS)$, $TE = E_i(x, y|VRS)$ and $SCE = E_i(x, y|CRS)/E_i(x, y|VRS)$. The first and second parts of Table 2 report this decomposition for the convex and nonconvex family of technologies. For each data set, there are three lines per efficiency component in a column: (i) the number of efficient observations, (ii) the average efficiency, and (iii) the Li (1996) test statistic. We comment on each of these three elements in turn.

For any efficiency component, it is well-known that the number of efficient observations is higher or equal under nonconvexity compared to the convex case. This number turns out to be equal for the OTE and SCE components in two data sets: Färe et al. (1985) and Färe et al. (1983). Average efficiency is also known to be higher or equal under nonconvexity, except for the SCE component since it is a ratio derived from the other two components.[7] This average turns out to be equal for the OTE component in just one data set: Färe et al. (1985).

One can assess the differences between convex and nonconvex efficiency estimates by using a test statistic initially proposed by

[7] The multiplicative decomposition of OTE need not hold exactly at the sample level, since arithmetic rather than geometric averages are reported.

Table 1
Sources of empirical data.

Article	Sample	# Inp.	# Outp.	Sector	Remarks
Färe et al. (1985)	32	3	1	Electricity	
Haag et al. (1992)	41	4	2	Agriculture	
Färe et al. (1983)	86	3	1	Electricity	Unbalanced (N= 20 and T = 5)
Cesaroni (2011)	92	2	5	Car registration	
Fan et al. (1996)	471	3	1	Agriculture	

Table 2
Decomposition of overall technical efficiency: convex and nonconvex.

Sample		Convexity			Nonconvexity			OTE^{NC}&
		OTE	*SCE*	*TE*	*OTE*	*SCE*	*TE*	$\neg OTE^C$
Färe et al. (1985)	#Eff. obs.	2	2	9	2	2	29	0
	Mean	0.905	0.952	0.951	0.905	0.906	0.998	0.000
	Li-test†				0.000	6.887***	8.533***	
Haag et al. (1992)	#Eff. obs.	8	8	10	20	20	41	12
	Mean	0.841	0.959	0.880	0.923	0.923	1.000	0.856
	Li-test†				4.208***	2.270**	NA	
Färe et al. (1983)	#Eff. obs.	4	4	18	4	4	68	0
	Mean	0.897	0.966	0.930	0.898	0.907	0.990	0.000
	Li-test†				0.000	21.926***	26.032***	
Cesaroni (2011)	#Eff. obs.	9	9	15	12	12	56	3
	Mean	0.652	0.876	0.733	0.702	0.761	0.911	0.972
	Li-test†				0.313	5.547***	21.123***	
Fan et al. (1996)	#Eff. obs.	18	18	49	60	60	164	42
	Mean	0.765	0.945	0.811	0.841	0.921	0.913	0.924
	Li-test†				18.459***	19.999***	52.878***	

† Li test: critical values at 1% level = 2.33 (***); 5% level = 1.64 (**); 10% level = 1.28 (*).

Li (1996) that is valid for both dependent and independent variables.[8] The null hypothesis of this Li-test states that both convex and nonconvex distributions for a given efficiency measure are equal. One can reject the null hypothesis of equal distributions for all components for the Fan et al. (1996) data set and for at least two components for all remaining data sets.[9] Thus, it seems rather safe to conclude that convex and nonconvex efficiency estimates differ for most components and data sets.

The last column of Table 2 reports both the number of CRS efficient observations under nonconvexity that are CRS inefficient under convexity, and the average amount of convexity-related *OTE* $(= E_i^C(x, y|CRS)/E_i^{NC}(x, y|CRS))$ for these same observations.[10] On the one hand, this is the net gain in the number of MPSS points due to dropping convexity. It varies between 0 and 42 observations among the data sets analyzed. On the other hand, convexity-related *OTE* indicates the amount of overall technical efficiency that can be attributed to the convexity axiom. Not surprisingly, this convexity-related *OTE* equals zero in two data sets: Färe et al. (1985) and Färe et al. (1983). On average, this amount varies between 0.856 and 0.972 when computed relative to the concerned observations: thus, the convex estimates suggest further gains in overall technical efficiency varying between 2.8% and 14.4%. Recall that Fig. 3 represents the section shown in both convex and nonconvex technologies depicted in Fig. 2: it clearly illustrates these cases where the nonconvex CRS technology is situated below the convex one. Thus, convex CRS technologies may well overestimate potential gains in average productivity.

Tables 3 and 4 each have two major parts. The first and second parts of Table 3 report on the percentage of observations relative to the sample size operating under increasing (IRS), constant (CRS) and decreasing returns to scale (DRS) for the convex and nonconvex technology, respectively. Table 4 again has two major parts. The first part lists the efficient observations on both technologies that share a common characterization of returns to scale for each of the three cases. The second part focuses on conflicting cases: switches from IRS to DRS (denoted IRS–DRS), from CRS to IRS (CRS–IRS), from CRS to DRS (CRS–DRS), and the total percentage of these conflicts relative to the sample size.

One can draw the following conclusions. First, the amount of common efficient observations spanning both technologies is quite modest. Obviously, the amount of common CRS observations is low because few observations are CRS efficient in the convex case in the first place. While the percentage of common IRS observations is low, especially the DRS part of technology is built on strikingly little common ground: almost no observations are in common. Second, apart from the first study with the smallest sample size, all other samples yield some minimal to moderate conflict in classification between convex and nonconvex technologies. This conflict varies from a modest about 7% for the Färe et al. (1983) sample to a quite substantial about 40% for the Haag et al. (1992) case, all three cases confounded. Third, the detailed sources of conflict in classification vary a lot among the different samples. While for the Haag et al. (1992) study the CRS–IRS conflict dominates for about 20% of observations, for Cesaroni (2011), Färe et al. (1983) and Fan et al. (1996) the IRS–DRS case is dominant: for a small about 7% for the first two cases to a substantial about 14% of observations for the third sample.

On the empirical evaluation of the incidence of the global sub-constant returns to scale case we can be very brief. We found none in any of the five samples investigated. This explains why this notion is not reported in any of the tables so far. It remains an open question which conditions determine the existence as well as the empirical incidence of this global sub-constant returns to scale case.

[8] Dependency is a basic characteristic of extremum or frontier estimators, since efficiency measures depend, among others, on sample size. Note that Fan and Ullah (1999) refine the same test.

[9] Note that this Li-test cannot be computed for the nonconvex TE component of Haag et al. (1992), since all observations are technically efficient and hence the kernel density cannot be estimated.

[10] The notion of convexity-related efficiency is introduced by Briec et al. (2004): for any input-oriented efficiency component it is the convex efficiency measure divided by the nonconvex one.

Table 3
Returns to scale on convex and nonconvex technologies: basic results.

Sample	Convexity			Nonconvexity		
	GIRS (%)	GCRS (%)	GDRS (%)	GIRS (%)	GCRS (%)	GDRS (%)
Färe et al. (1985)	78.13	6.25	15.63	78.13	6.25	15.63
Haag et al. (1992)	53.66	19.51	26.83	43.90	48.78	7.32
Färe et al. (1983)	66.28	4.65	29.07	73.26	4.65	22.09
Cesaroni (2011)	83.70	9.78	6.52	78.26	13.04	8.70
Fan et al. (1996)	52.44	3.82	43.74	52.23	12.74	35.03

Table 4
Returns to scale on convex and nonconvex technologies: common efficient observations and conflicts.

Sample	# Common effic. obs.			Conflicting cases			
	GIRS (%)	GCRS (%)	GDRS (%)	GIRS–GDRS (%)	GCRS–GIRS (%)	GCRS–GDRS (%)	Total conflicts (%)
Färe et al. (1985)	15.63	6.25	6.25	0.00	0.00	0.00	0.00
Haag et al. (1992)	4.88	19.51	0.00	9.76	19.51	9.76	39.02
Färe et al. (1983)	16.28	4.65	0.00	6.98	0.00	0.00	6.98
Cesaroni (2011)	2.17	9.78	1.09	7.61	2.17	1.09	10.87
Fan et al. (1996)	1.70	3.82	1.91	13.80	3.40	5.52	22.72

4.3. The behavior of RAP

To illustrate the role played by the behavior of RAP in the above results to scale, and how it may affect local returns to scale, we have chosen to depict some typical observations selected from the Cesaroni (2011) sample. In particular, we have selected observation 40 because it is efficient in both the convex and nonconvex CRS technologies. Then, we depict observation 44 representing the conflict between GIRS–GDRS.[11]

For each of these cases we show a pair of figures: the above represents the optimal (δ, γ)-combinations of a section from the origin through the observation in input–output space; the below depicts the evolution of RAP along the same radial section. The observations under scrutiny are situated at the coordinates (1, 1) in both the upper and lower parts of the figures. The convex (nonconvex) case is shown as a dashed (continuous) line. Note that for the RAP figure one must distinguish between points where RAP is smaller and larger than unity: only the latter points indicate improvements with respect to the observation under evaluation and are candidates for optima. RAP points smaller than unity form at best a local optimum or sub-optimum for themselves in that RAP may be stationary at such points. But, these points can never be optimal since the RAP level is below that of the observation under examination.

We first comment on observation 40 depicted in Fig. 4. Being a unique optimal MPSS point labeled *A*, there is a close to optimal point labeled *B* to the right where RAP is close to constant under convexity but varies a lot under nonconvexity. Beyond this point *B* to the right RAP declines monotonously under convexity and more rapidly and close to monotonously except for the end of the empirical range under nonconvexity. To the left of the MPSS point *A*, RAP declines monotonously, albeit more rapidly again under nonconvexity.

Next, we comment on observation 44 shown in Fig. 5. As can be noticed from the upper part, this observation is situated under DRS (IRS) under convexity (nonconvexity). In the lower part, it is clearly visible that the MPSS point under nonconvexity labeled *A* is situated to the right of unity, while the MPSS point under convexity labeled *B* is positioned slightly to the left of unity and suggests a higher RAP than the nonconvex case. This is a perfect illustration of the phenomenon depicted by a numerical example in Fig. 3. To both the right and especially the left from the nonconvex MPSS point *A*, there is quite some variation: there are three local optima of RAP to the right and at least six local optima of RAP to its left. Under convexity, the RAP curve suggests a smooth rise and decline around its optimal point *B*.

Table 5
Local returns to scale in the nonconvex technology.

Sample	# Units				
	Local IRS	Local CRS	Local DRS	Local SCRS	Total
Cesaroni (2011)	9	78	5	0	92

Summarizing these results, one can draw two preliminary conclusions. First, the evolution of RAP under nonconvexity is not smooth at all and reveals a variety of local optima that remain hidden in the smooth increase and subsequently decrease of RAP in the convex case. Remedying issues of suboptimal scale size is rather straightforward under convexity. Any diagnosis of IRS or DRS leads to an unambiguous recommendation to either increase or decrease the scale of operations, whereby any step in the right direction monotonously increases or decreases RAP, respectively. Under nonconvexity remedying the scale of operations is much harder and depends on choosing the right step size to either increase or decrease the scaling of the unit under evaluation. In empirical applications, there seem to be many areas where the lack of data is filled up by the convexity axiom, while the nonconvex approach clearly reveal the gaps in the empirical range of the data and our ensuing lack of knowledge about the technology. Of course, it cannot be excluded that sector specialists (managers, engineers, regulators, etc.) may have an a priori understanding on which ranges of operation are actually feasible even though these are currently not supported by the empirical range of the data.

Second, under nonconvexity several relative maxima of RAP could in principle occur even in a significantly smaller range than that considered in Fig. 5, thus affecting local returns to scale results.

4.4. Local returns to scale

Hereafter, we present the results deriving from the application of the FDH local method proposed in Definition 3.2 to the Cesaroni (2011) sample. Table 5 illustrates the classification associated to the interval-size $\varepsilon = 0.025$, which has been computed by means of the application of Soleimani-damaneh et al. (2006)

[11] These observation numbers are internal numbers attributed by us. We are not in a position to disclose the identity of these observations.

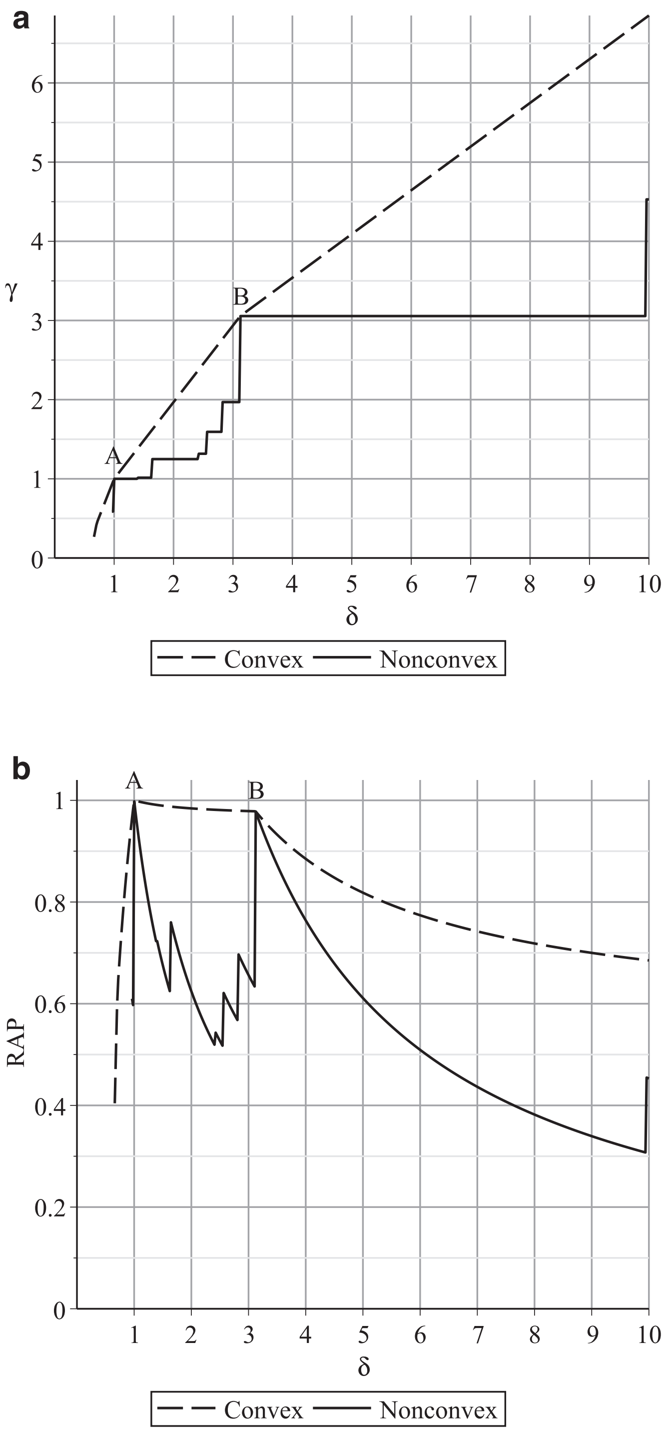

Fig. 4. Representation of radial section for observation 40 in input–output space (a) and evolution of RAP (b).

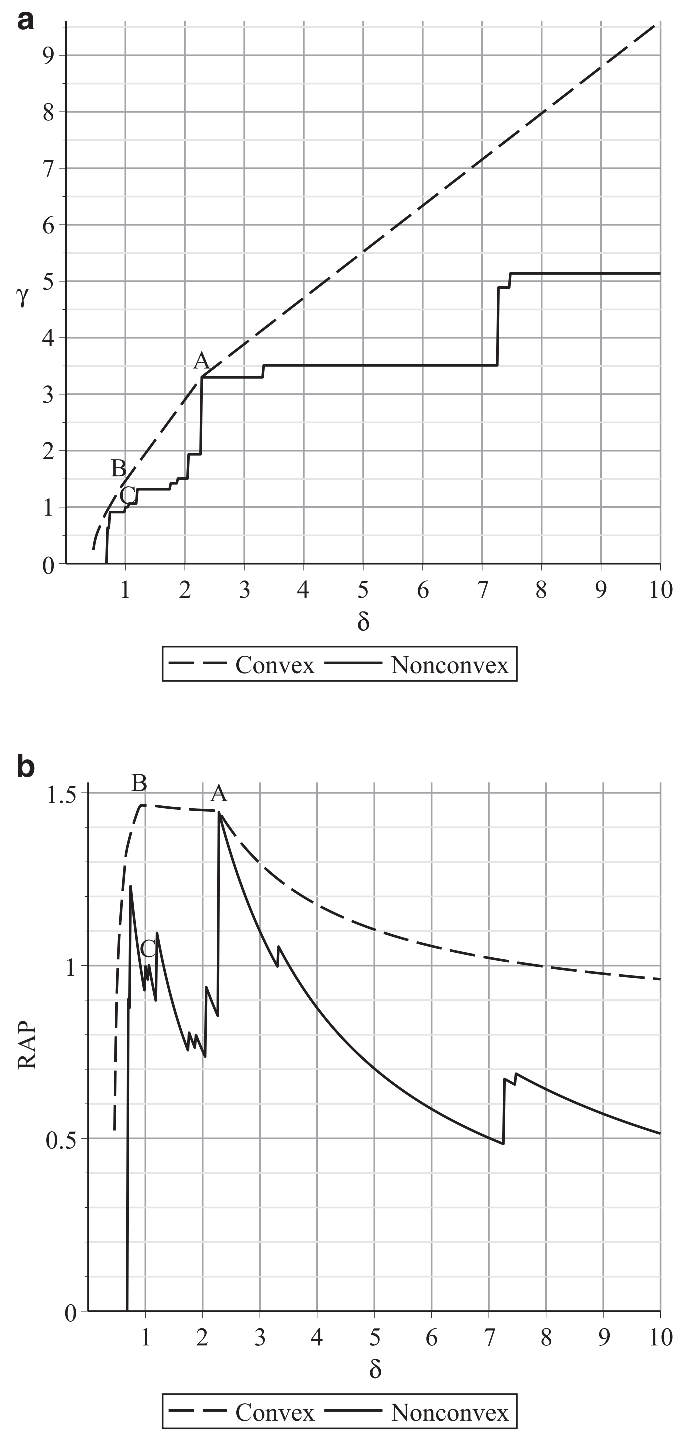

Fig. 5. Representation of radial section for observation 44 in input–output space (a) and evolution of RAP (b).

enumeration-algorithm to frontier points contained in the 0.025 neighborhood of each efficient point under examination.

We found no local sub-constant returns to scale case while the majority of observations exhibits local CRS. Note however that 14 units leave the default CRS-characterization in favor of either the IRS (9 units) or the DRS (5 units) condition. This fact indicates that relative maxima of RAP are actually present in the 0.025 neighborhood of these DMUs. Figs. 6 and 7 illustrate a local DRS (observation 45) and IRS case (observation 83), respectively. Observation 45 clearly has just one relative maximum of RAP situated at the left end of the neighborhood.

The case of observation 83 is particularly interesting due to the replication in a small interval of the complex behavior of RAP we have illustrated for the global case. In fact, in the increasing direction, we can observe the occurrence of three relative maxima of RAP within a 0.008 range with the greatest maximum delivering a substantial increase in RAP (about 50%). This utterly clarifies the remark we made at the end of Section 3.2 regarding one advantage of our local classification method.

These local returns to scale results are hard to summarize neatly using standard descriptive statistics. While their local nature lends itself excellently for depiction, this approach carries the risk

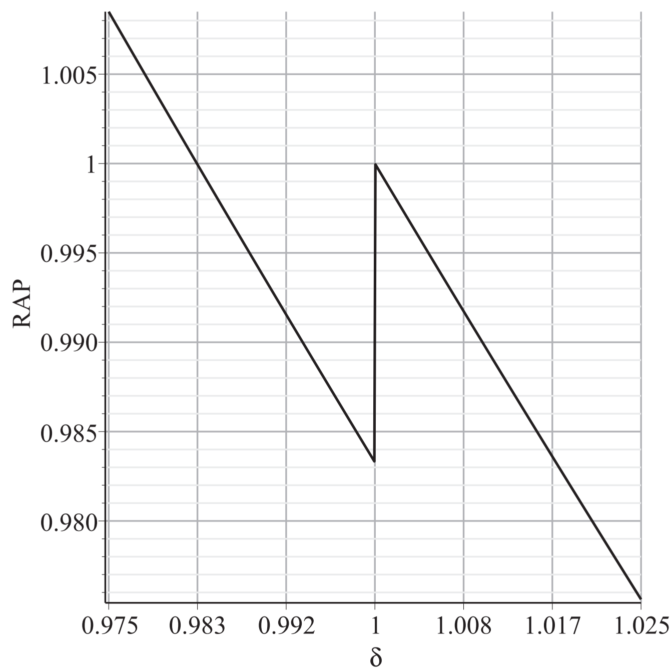

Fig. 6. Local evolution of RAP for observation 45.

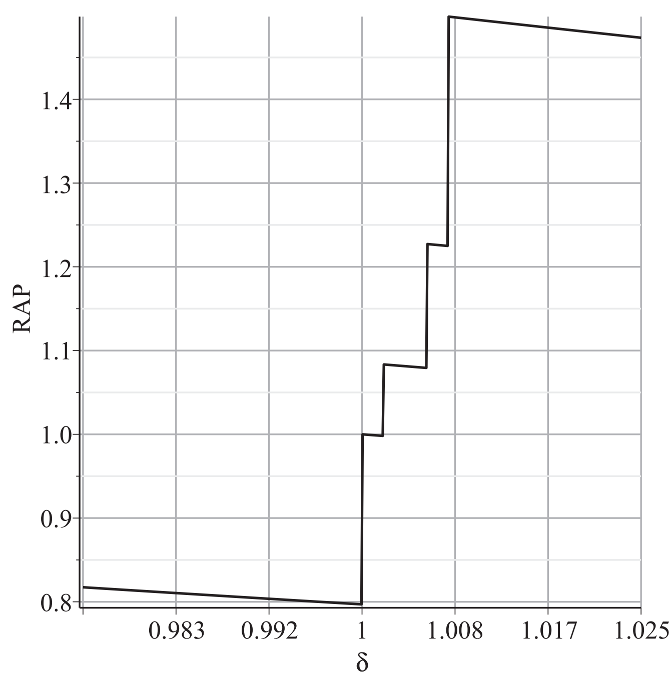

Fig. 7. Local evolution of RAP for observation 83.

that an empirical analysis becomes somehow casuistic and does not allow to draw any general conclusions at the sample level. Lacking standards to report the results of nonconvex analysis in economics, this problem cannot be easily solved in the short run.

5. Conclusions

Starting from the seminal contributions of Podinovski (2004a, 2004b) who characterizes both the notions of global and local returns to scale for nonconvex technologies, this contribution leads to three main conclusions.

First, we have clearly empirically established that the characterization of returns to scale on convex and nonconvex technologies may yield conflicting advice for substantial parts of samples. This confirms that Podinovski (2004a, 2004b) was certainly right in further scrutinizing the notion of returns to scale for nonconvex nonparametric technologies.

Second, while Podinovski (2004a, 2004b) convincingly argued for the existence a fourth type of global sub-constant returns to scale case complementing the three traditional cases (constant, decreasing and increasing returns to scale), our empirical tests reveal that none of the five secondary data sets analyzed contains a single observation that experiences such global sub-constant returns to scale. Which conditions determine the existence as well as the empirical incidence of this global sub-constant returns to scale case remains a question for future research.

Third, we have made a start to explore the differences between global and local returns to scale characterizations on FDH models. Especially the local results are revealing in that these clearly show how RAP evolves nonsmoothly and nonmonotonously under nonconvexity, while it is smooth and monotonous for convex nonparametric technologies. As spelled out earlier, this makes remedying scale deficiencies much harder under nonconvexity.

To the best of our knowledge, this is the first contribution that has managed to shed some light on all these issues. Of course, much more remains to be done. For instance, outliers are an issue for all nonparametric technology specifications and it could be interesting to evaluate how these affect the empirical differences as to returns to scale observed between convex and nonconvex technologies. As another example, a comparison among other definitions of local returns to scale could prove insightful when analyzing nonconvex nonparametric technologies. Finally, while this research has been confined to analyzing changes along a radial section in input–output space, keeping in mind that some managers may well prefer mimicking actual observations (e.g., Halme, Korhonen, & Eskelinen, 2014), it could be interesting to also develop an average productivity notion along a non-radial rather than a radial path.[12]

Appendix A. Note on equivalence of RTS and direction to MPSS in any convex technology

The equivalence between local and global returns-to-scale characterizations in convex production technologies is pointed out by Podinovski (2004a, p. 228): “This dual role of RTS classification as a local improvement indicator and direction to MPSS is preserved in any convex technology”. In a smooth production technology, this conclusion is established by means of Theorem 7 in Podinovski (2004a, p. 249). In a polyhedral production technology, the same property is due to the results that Banker (1984) and Banker and Thrall (1992) obtain by means of the comparison between the constant returns to scale (CCR) and the variable returns to scale (BCC) models, as we are going to illustrate in the remainder. In brief, this equivalence holds for standard (i.e., one-stage) convex production technologies.

CRS case:

In a CRS technology, by definition, an efficient point $(x_0, y_0) > 0$ is an optimal solution to the CCR problem (2) in Banker and Thrall (1992, p. 77) with $r_0^* = 1$, then according to their Proposition 1 this point is an MPSS. Given the fact that in the CRS technology this efficient point is not an extreme scale size, Proposition 1 in Banker (1984) can then be used to show that local constant returns-to-scale prevail at this MPSS.

[12] See, e.g., Chambers and Mitchell (2001) for other examples on the importance of nonradial changes.

IRS and DRS cases:

The equivalence results achieved by Banker and Thrall (1992) for the VRS technology (BCC) are quite general, because referred to an efficient frontier which allows for different portions, each satisfying different RTS assumption (CRS, DRS or IRS). The general validity of their Propositions 3 and 4 can be illustrated by means of a simple graphical example regarding the IRS case. Mutatis mutandis, the same conclusion will symmetrically hold for a DRS frontier.

Consider Fig. 9 in Podinovski (2004a, p. 250) as a representation of a convex technology with an IRS efficient-frontier. Each of the frontier points comprised between A and C exhibits local increasing returns to scale according to Definition 8 and Proposition 3 (Banker and Thrall, 1992, p. 79). These points are not extreme scale sizes and therefore Proposition 4 (see Banker and Thrall, 1992, p. 80) can be applied to conclude that these points are less than their MPSS (which is in fact represented by point C). As far as point A is concerned, which is a smallest scale size, we note that according to Definition 8-footnote 5 (see Banker and Thrall, 1992, p. 79) it can be classified as showing local increasing returns to scale, because $\rho^* > 1$; here, Proposition 4 cannot be directly applied, but it is nevertheless evident that A is less than its MPSS (point C). Finally, with regard to C we note that, according to Definition 8, it exhibits constant returns to scale and that – given the fact that it is not an extreme scale size – Proposition 4 holds.

References

Afriat, S. (1972). Efficiency estimation of production functions. *International Economic Review, 13*(3), 568–598.
Alam, I., & Sickles, R. (1998). The relationship between stock market returns and technical efficiency innovations: Evidence from the US airline industry. *Journal of Productivity Analysis, 9*(1), 35–51.
Balaguer-Coll, M., Prior, D., & Tortosa-Ausina, E. (2007). On the determinants of local government performance: A two-stage nonparametric approach. *European Economic Review, 51*(2), 425–451.
Banker, R. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research, 17*(1), 35–44.
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.
Banker, R., Cooper, W., Seiford, L., Thrall, R., & Zhu, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research, 154*(2), 345–362.
Banker, R., & Thrall, R. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research, 62*(1), 74–84.
Briec, W., Kerstens, K., Leleu, H., & Vanden Eeckaut, P. (2000). Returns to scale on nonparametric deterministic technologies: Simplifying goodness-of-fit methods using operations on technologies. *Journal of Productivity Analysis, 14*(3), 267–274.
Briec, W., Kerstens, K., & Vanden Eeckaut, P. (2004). Non-convex technologies and cost functions: Definitions, duality and nonparametric tests of convexity. *Journal of Economics, 81*(2), 155–192.
Cesaroni, G. (2011). A complete FDH efficiency analysis of a diffused production network: The case of the Italian driver and vehicle agency. *International Transactions in Operational Research, 18*(2), 205–229.
Cesaroni, G., & Giovannola, D. (2015). Average-cost efficiency and optimal scale sizes in non-parametric analysis. *European Journal of Operational Research, 242*(1), 121–133.
Chambers, R., & Mitchell, T. (2001). Homotheticity and non-radial changes. *Journal of Productivity Analysis, 15*(1), 31–39.
Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
Cummins, D., & Zi, H. (1998). Comparison of frontier efficiency methods: An application to the U.S. life insurance industry. *Journal of Productivity Analysis, 10*(2), 131–152.
Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor efficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurements* (pp. 243–268). Amsterdam: North Holland.
Destefanis, S. (2003). The Verdoorn law: Some evidence from non-parametric frontier analysis. In J. McCombie, M. Pugno, & B. Soro (Eds.), *Productivity growth and economic performance: Essays on Verdoorn's law* (pp. 136–164). Basingstoke: Palgrave Macmillan.
Diewert, W., & Parkan, C. (1983). Linear programming test of regularity conditions for production functions. In W. Eichhorn, K. Neumann, & R. Shephard (Eds.), *Quantitative studies on production and prices* (pp. 131–158). Würzburg: Physica-Verlag.
Emrouznejad, A., & Amin, G. (2009). DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling, 33*(1), 486–498.
Emrouznejad, A., Parker, B., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences, 42*(3), 151–157.
Evanoff, D., & Israilevich, P. (1995). Scale elasticity versus scale efficiency in banking. *Southern Economic Journal, 61*(4), 1036–1046.
Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics, 14*(4), 460–468.
Fan, Y., & Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics, 11*(1), 337–360.
Färe, R., Grosskopf, S., & Logan, J. (1983). The relative efficiency of Illinois electric utilities. *Resources and Energy, 5*(4), 349–367.
Färe, R., Grosskopf, S., Logan, J., & Lovell, C. (1985). Measuring efficiency in production: With an application to electric utilities. In A. Dogramaci, & N. Adam (Eds.), *Managerial issues in productivity analysis* (pp. 185–214). Boston: Kluwer.
Färe, R., Grosskopf, S., & Lovell, C. (1983). The structure of technical efficiency. *Scandinavian Journal of Economics, 85*(2), 181–190.
Førsund, F., & Hjalmarsson, L. (2004). Calculating scale elasticity in DEA models. *Journal of the Operational Research Society, 55*(10), 1023–1038.
Førsund, F., Hjalmarsson, L., Krivonozhko, V., & Utkin, O. (2007). Calculation of scale elasticities in DEA models: Direct and indirect approaches. *Journal of Productivity Analysis, 28*(1–2), 45–56.
Haag, S., Jaska, P., & Semple, J. (1992). Assessing the relative efficiency of agricultural production units in the blackland prairie, Texas. *Applied Economics, 24*(5), 559–565.
Hackman, S. (2008). *Production economics: Integrating the microeconomic and engineering perspectives*. Berlin: Springer.
Hadjicostas, P., & Soteriou, A. (2006). One-sided elasticities and technical efficiency in multi-output production: A theoretical framework. *European Journal of Operational Research, 168*(2), 425–449.
Halme, M., Korhonen, P., & Eskelinen, J. (2014). Non-convex value efficiency analysis and its application to bank branch sales evaluation. *Omega, 48*, 10–18.
Kerstens, K., & Vanden Eeckaut, P. (1999). Estimating returns to scale using nonparametric deterministic technologies: A new method based on goodness-of-fit. *European Journal of Operational Research, 113*(1), 206–214.
Kerstens, K., & Van de Woestyne, I. (2014). Solution methods for nonconvex free disposal hull models: A review and some critical comments. *Asia-Pacific Journal of Operational Research, 31*(1), 1450010-1–1450010-13.
Krivonozhko, V., Utkin, O., Volodin, A., Sablin, I., & Patrin, M. (2004). Constructions of economic functions and calculations of marginal rates in DEA using parametric optimization methods. *Journal of the Operational Research Society, 55*(10), 1049–1058.
Leleu, H. (2006). A linear programming framework for free disposal hull technologies and cost functions: Primal and dual models. *European Journal of Operational Research, 168*(2), 340–344.
Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews, 15*(1), 261–274.
Podinovski, V. (2004a). Efficiency and global scale characteristics on the "no free lunch" assumption only. *Journal of Productivity Analysis, 22*(3), 227–257.
Podinovski, V. (2004b). Local and global returns to scale in performance measurement. *Journal of the Operational Research Society, 55*(2), 170–178.
Podinovski, V. (2004c). On the linearisation of reference technologies for testing returns to scale in FDH models. *European Journal of Operational Research, 152*(3), 800–802.
Ray, S. (2004). *Data envelopment analysis: Theory and techniques for economics and operations research*. Cambridge: Cambridge University Press.
Sahoo, B., Zhu, J., Tone, K., & Klemen, B. (2014). Decomposing technical efficiency and scale elasticity in two-stage network DEA. *European Journal of Operational Research, 233*(3), 584–594.
Scarf, H. (1981a). Production sets with indivisibilities part i: Generalities. *Econometrica, 49*(1), 1–32.
Scarf, H. (1981b). Production sets with indivisibilities part ii: The case of two activities. *Econometrica, 49*(2), 395–423.
Seiford, L. (1997). A bibliography for data envelopment analysis (1978–1996). *Annals of Operations Research, 73*, 393–438.
Seiford, M., & Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega, 27*(1), 1–11.
Shephard, R. (1967). The notion of a production function. *Mathematical Methods of Operations Research, 11*(1), 209–232.
Soleimani-damaneh, M., Jahanshahloo, G., & Reshadi, M. (2006). On the estimation of returns-to-scale in FDH models. *European Journal of Operational Research, 174*(2), 1055–1059.
Soleimani-damaneh, M., & Mostafaee, A. (2009). Stability of the classification of returns to scale in FDH models. *European Journal of Operational Research, 196*(3), 1223–1228.
Soleimani-damaneh, M., & Reshadi, M. (2007). A polynomial-time algorithm to estimate returns to scale in FDH models. *Computers and Operations Research, 34*(7), 2168–2176.
Tone, K., & Sahoo, B. (2003). Scale, indivisibilities and production function in data envelopment analysis. *International Journal of Production Economics, 84*(2), 165–192.
Varian, H. (1984). The nonparametric approach to production analysis. *Econometrica, 52*(3), 579–597.