Document de travail du LEM / Discussion paper LEM
2019-13

Ray Average Cost in Convex Technologies: Some Generalisations

Giovanni CESARONI
Department for public administration, Prime Minister's Office / g.cesaroni@governo.it

Kristiaan KERSTENS
LEM UMR 9221 / k.kerstens@ieseg.fr

Ignace VAN de WOESTYNE
KU Leuven, Research unit MEES / ignace.vandewoestyne@kuleuven.be

http://lem.cnrs.fr/

Ray Average Cost in Convex Technologies:

Some Generalisations

Giovanni Cesaroni†

Kristiaan Kerstens‡

Ignace Van de Woestyne§

We propose a model for the estimation of scale economies (cost-based returns to scale analysis) based on the minimization of the ray average cost function. It provides both a simplified computational approach and more relevant economic information when compared to existing methods. In particular, for the frontier points under evaluation, we can avoid both the need to solve the non-increasing returns to scale problem, required in Färe and Grosskopf (1985), and that of solving the complex dual-problem set by Sueyoshi (1999) to compute the intercept of the supporting hyperplane and the corresponding value of the cost-based scale elasticity. By both allowing for multiple solutions to the scale efficiency problem and determining the convex shape of the ray average cost function in the outputs, our analysis also proves the impossibility of global sub-constant scale economies and the equivalence between local and global classifications. Furthermore, we show that the RAC function is convex in the outputs if and only if the technology is convex. These three issues have not been dealt with by the existent literature on cost models.

Keywords: Returns to scale, Scale economies, Ray average cost

7 May 2019

† Department for public administration, Prime Minister's Office, Via del Sudario 49, I-00186 Rome, Italy. g.cesaroni@governo.it

‡ IÉSEG School of Management, CNRS-LEM (UMR 9221), Univ. Lille, 3 rue de la Digue, FR-59000 Lille, France. k.kerstens@ieseg.fr Corresponding author.

§ KU Leuven, Research unit MEES, Warmoesberg 26, BE-1000 Brussels, Belgium. ignace.vandewoestyne@kuleuven.be

1. Introduction

The introduction of the concept of ray average cost (RAC) as a tool to investigate scale economies in multiple output technologies dates back to the seminal contributions of Panzar and Willig (1977) and Baumol et al. (1982). However, these authors ignore the possibility of inefficiency in the production technology. Chavas and Cox (1999, p. 308) are the first to define the RAC function in the context of production technologies with inefficiency and to point out that the concept is strictly connected to the evaluation of allocative inefficiency *via* the constant returns-to-scale (CRS) cost function. In the remainder of this contribution we concentrate on technologies and cost functions allowing for inefficiencies. For simplicity, our arguments are developed in a non-parametric framework. However, our key results are perfectly general.

Despite these developments, the measurement of scale economies in non-parametric production models has -relatively speaking- attracted little analytical interest compared to the vast literature on production-based returns to scale (see, e.g., Banker, 1984; Banker and Thrall, 1992; Banker et al. 1996, Färe et al. 1983). The latter contributions have in fact dealt with the maximization of ray average productivity of a decision-making unit (DMU) while ignoring the possible allocative inefficiency of its input mix. In this context, Tone and Sahoo's (2003) remark that 'scale in all its definitions warrants the input mix and output mix to remain constant' remains relevant. As a matter of fact, also more recent articles (like those of Podinovski 2004, Førsund and Hjalmarsson 2004, Førsund et al. 2009) still rely exclusively on the maximization of ray average productivity.

For convex technologies, the available methods for the estimation of cost-based returns to scale with a variable input mix are at present limited to those developed by Färe and Grosskopf (1985) and Sueyoshi (1999). On the one hand, Färe and Grosskopf (1985) use scale efficiency to determine global qualitative information on the prevailing regime of scale economies. In particular, these authors solve three different programming problems under variable (VRS), constant (CRS) and non-increasing returns to scale (NIRS). However, this method neglects the possible occurrence of global sub-constant scale economies (GSCSE) due to the existence of multiple optimal solutions (see Podinovski (2004)), i.e., the case where the same level of the CRS cost can be achieved by either increasing or decreasing the current scale size of a cost-scale inefficient unit (see Cesaroni et al. 2017b, p. 1443). On the other hand, Sueyoshi (1999) introduces a cost-based scale elasticity to deliver quantitative information on the precise degree of scale economies. In this case, multiple solutions to the CRS problem are not considered, while the proposed method is purely local.

Therefore, we may conclude that none of these two modeling strategies addresses the two issues represented by the (i) multiplicity of the CRS solutions to the scale-efficiency problem and (ii) the relation between global and local classifications of scale economies in a VRS technology. By contrast, for convex production-based returns to scale these analytical problems have been solved in Banker (1984), Banker and Thrall (1992) and Podinovski (2004). In particular, Podinovski (2004) determines explicitly the behavior of the ray average productivity (RAP) function of a DMU over the entire domain, thereby confirming the coincidence of global and local returns-to-scale classifications (for a more detailed discussion and an empirical exploration: see Cesaroni et al., 2017a).

Unfortunately, precisely due to the potential allocative inefficiency of the input mix, Cesaroni and Giovannola (2015) have shown that these RAP results are in principle not applicable to RAC analysis, because of the non-coincidence of the most productive and optimal scale sizes of a DMU based on the maximization of RAP and minimization of RAC respectively. Therefore, a specific analysis is needed to establish whether the GSCSE case and the non-monotonic behavior of RAC, established by Cesaroni et al. (2017b) in a non-convex setting, may also occur in convex technologies. With regard to this second aspect, we also note that -to the best of our knowledge- the literature on non-parametric production analysis has so far failed to supply a general characterization of the RAC function under convexity (cfr., e.g., Chavas and Cox 1999, Førsund and Hjalmarsson 2004, Ray 2015).

Therefore, this contribution sets itself four goals for the specific case of convex technologies. First, we show that the RAC function is convex in the outputs if and only if the technology is convex. Otherwise, it is non-convex. This is akin to the often-neglected property of the cost function that it is convex in the outputs if and only if the technology is convex (e.g., Shephard (1970: p.227)). Note that while this result is perfectly general, we develop it using non-parametric production and cost function models without invoking any differentiability assumption. Second, we determine the behavior of the RAC function of a DMU over the entire domain. This allows exploring the relation between the global and the local classification of scale economies and we introduce a new local measure of the degree of scale economies, as an alternative to the one by Sueyoshi (1999). Third, by following the non-parametric Banker (1984) approach and building upon the RAC, we propose a new simplified procedure for the evaluation of global scale economies when multiple CRS solutions are possible and quantitative information on the degree of scale economies is provided. Fourth, we empirically illustrate the use of our new method in comparison with those of Färe and Grosskopf (1985) and Sueyoshi (1999).

The contribution is structured as follows. Section 2 introduces some preliminary definitions and briefly discusses the classification problems that can in principle affect the two prevailing methods of Färe and Grosskopf (1985) and Sueyoshi (1999). Section 3 illustrates the main features of the RAC classification method and focuses on the multiplicity of CRS solutions and the eventual possibility of the GSCSE case. Section 4 deals with the behavior of the RAC of a DMU's output mix in an expansion/contraction to an optimal scale size, which provides information on the relationship between global and local scale economies. Section 5 uses the proposed method to estimate global and local scale economies on two multiple inputs and outputs datasets, while comparing our results with those obtained from the competing methods. Furthermore, we briefly illustrate the difference between RAP and RAC functions. Section 6 summarizes our contribution to the literature.

2. Current Approaches to Evaluating Scale Economies

2.1 Preliminary Definitions and Review

The non-parametric production technology considered is the convex hull of the observed production possibilities (see, e.g., Banker et al. 1984, Färe and Grosskopf 1985). Introducing the notation, we have n observations indexed by j $(j = 1,...,n)$ using m inputs x_{ij} $(i = 1,...,m)$ that are capable to produce s outputs y_{rj} $(r = 1,...,s)$. The observed input and output vectors are $\mathbf{x}_j = (x_{1j},...,x_{mj})' \geq \mathbf{0}$ and $\mathbf{y}_j = (y_{1j},...,y_{sj})' \geq \mathbf{0}$, respectively, with $\mathbf{x}_j, \mathbf{y}_j \neq \mathbf{0}$. Note that the prime denotes the transposition operation. Denoting the $m \times n$ matrix of inputs as $\mathbf{X} = [\mathbf{x}_1,..., \mathbf{x}_n]$ and the $s \times n$ matrix of outputs as $\mathbf{Y} = [\mathbf{y}_1,..., \mathbf{y}_n]$, then the production possibility set or technology can be expressed as

$$T_K = \left\{ (\mathbf{x}, \mathbf{y}) \middle| \, \mathbf{Xz} \leq \mathbf{x}, \mathbf{Yz} \geq \mathbf{y}, z_j \equiv w\lambda_j, \lambda_j \geq 0, \sum_{j=1}^{n} \lambda_j = 1 \right\}, \tag{1}$$

where $\mathbf{z}$ is the $n \times 1$ vector with components equal to $w\lambda_j$. The scalar $w > 0$ is a scaling factor introducing different returns to scale assumptions in the technology indexed by *K*. These scaling options are the following: $w > 0 \Leftrightarrow T_{CRS}, w = 1 \Leftrightarrow T_{VRS}, w \leq 1 \Leftrightarrow T_{NIRS}, w \geq 1 \Leftrightarrow T_{NDRS}$, whereby NDRS stands for non-decreasing returns to scale. Note that T_{VRS} is the least restrictive technology, since it generates the smallest set enveloping all original observations.

Turning to the analysis of costs, we denote the vector of observed input prices as $\mathbf{p}_j = (p_{1j},..., p_{mj}) > \mathbf{0}$ with $\mathbf{p}_j\mathbf{x}_j$ representing the total observed cost of observation j for producing its output vector $\mathbf{y}_j$, which also represents its current scale size.[1] In any of the K technologies defined above, the cost efficiency of a given unit requires solving the following programming problem:

$$\begin{aligned}&\underset{\mathbf{z}}{\mathrm{Min}}\ \frac{\mathbf{p}_j\mathbf{X}\mathbf{z}}{\mathbf{p}_j\mathbf{x}_j}\\ &\text{s.t.}\\ &\mathbf{Y}\mathbf{z}\geq\mathbf{y}_j\end{aligned}\qquad(2)$$

An optimal solution to program (2) $\mathbf{z}^*$ determines the cost efficiency score of observation j as $\frac{\mathbf{p}_j\mathbf{X}\mathbf{z}^*}{\mathbf{p}_j\mathbf{x}_j}$. This ratio of minimal to observed costs equals unity for a cost-efficient DMU.

We end this subsection with another definition.

Definition 1 For an observation j global sub-constant scale economies occur when in T_{CRS}, $\frac{\mathbf{p}_j\mathbf{X}\mathbf{z}^*}{\mathbf{p}_j\mathbf{x}_j}<1$ and there are at least two solutions $\mathbf{z}_1^*$ and $\mathbf{z}_2^*$ such that $\mathbf{p}_j\mathbf{X}\mathbf{z}_1^*=\mathbf{p}_j\mathbf{X}\mathbf{z}_2^*=\mathbf{p}_j\mathbf{X}\mathbf{z}^*$ with $\sum_{j=1}^{n}z_{j1}^*<1$ and $\sum_{j=1}^{n}z_{j2}^*>1$.

We remark that global sub-constant scale economies can be interpreted as the case where a scale-inefficient unit has the same CRS optimal-cost determined by two different scale sizes $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ belonging to the same VRS technology, one of which is larger and the other one which is smaller than the current scale size. Mathematically, this reads: $\mathbf{Y}\mathbf{z}_{\mathbf{1}}^*\equiv w_1^*\mathbf{y}_1^*$ and $\mathbf{Y}\mathbf{z}_2^*\equiv w_2^*\mathbf{y}_2^*$ respectively, with $w_1^*<1$, $w_2^*>1$, $w_1^*\mathbf{y}_1^*\geq\mathbf{y}_j$, $w_2^*\mathbf{y}_2^*\geq\mathbf{y}_j$, and where w is the scaling factor from (1).

[1] A scale size variable indicates the level at which either inputs or outputs are actually being employed by a unit under evaluation (i.e., current). In the analysis of scale economies, due to the possibility of large variations in input proportions, this scale size variable is typically expressed in terms of the outputs. We follow this convention.

2.2 Classification Issues in the Färe and Grosskopf (1985) and Sueyoshi (1999) Models

The method of Färe and Grosskopf (1985: p. 600) computes the cost scale efficiency defined as the ratio between CRS and VRS cost efficiency scores of the points lying on the frontier of the technology. If the point under examination is cost-scale efficient, then it displays CRS. Otherwise, a third NIRS technology is employed to establish qualitative information on the nature of scale inefficiency, i.e., on the direction to the VRS optimal scale size whose projection determines the CRS cost efficiency score. To be precise, when NIRS and VRS cost efficiencies are identical, then scale inefficiency is due to decreasing returns to scale (DRS): i.e., the optimal scale size is lower than the current one. Otherwise, when NIRS and VRS cost efficiencies do not coincide, then scale inefficiency is due to increasing returns to scale (IRS): i.e., the optimal scale size is greater than the current one. This method ignores the possibility of multiple CRS solutions. Following Podinovski (2004), we can denote the classification in question as global because it is based on the absolute-minimum cost (i.e., the CRS cost) and it is determined by a scale size which may be rather distant from the current scale examined. Note that this method does not provide quantitative information relating to the degree of scale economies, commonly measured by the scale elasticity (i.e., the ratio of average to marginal cost).

This scale elasticity approach is developed by Sueyoshi (1999) to ascertain the scale economies regime of frontier points. In addition to a VRS cost-efficiency problem (2), a complex programming problem, which is the dual of the former (see Sueyoshi 1999, pp. 1599-1601), is set to compute the cost-scale elasticity e_c at the projection of an observation onto the efficient frontier. In analogy with the returns-to-scale analysis of Banker et al. (1984) and Banker and Thrall (1992), this elasticity can be interpreted as the intercept of the supporting hyperplane in the cost-outputs space. The cost-efficient point under examination exhibits increasing, decreasing and constant scale economies if, respectively, $e_c > 1$, $e_c < 1$, $1 \le e_c \le 1$ hold (see Sueyoshi 1999, p. 1603). In particular, we may remark that this method does not discuss the relationship between a solution to the CRS cost-efficiency problem and a VRS solution featuring local constant scale economies, while also multiple solutions to the CRS problem are not taken into account (cfr., *a contrario*, Banker 1984, Banker and Thrall 1992 in a returns-to-scale setting).[2]

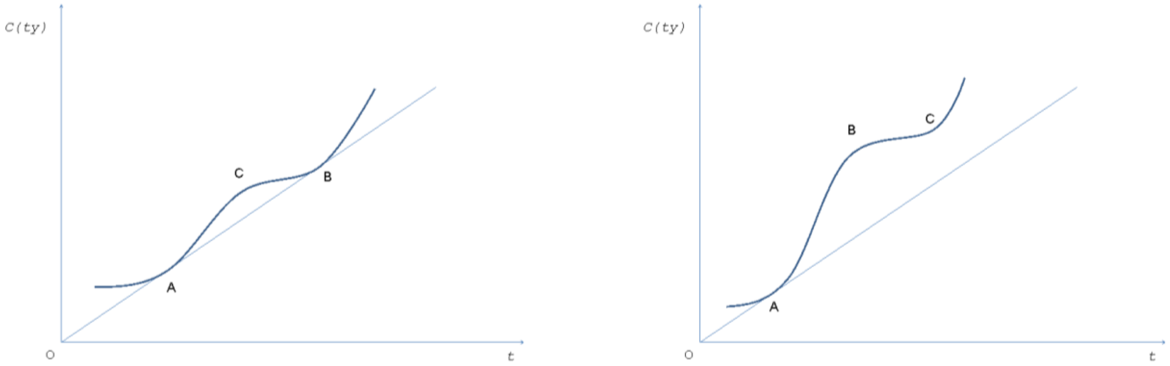
Complementing the RAP analysis of Podinovski (2004), Cesaroni et al. (2017b) have shown that in a non-convex technology the multiplicity of CRS solutions and the non-monotonicity of the RAC function may lead to both an erroneous global classification and a non-coincidence of global and local scale economies classifications. These problems are illustrated in Figure 1, where for the ease of exposition a smooth technology is assumed. For given input prices, the cost function $C(ty)$ is represented on the vertical axis. The radial expansion/contraction factor of the output mix of an evaluated DMU, $t > 0$, is represented along the horizontal axis. The RAC is defined as the ratio $C(ty)/t$.

Figure 1 (left) describes the hypothetical case of multiple CRS solutions and the non-monotonic behavior of RAC. The Färe and Grosskopf (1985) criterion classifies point C as global increasing scale economies, while it is in fact characterized by GSCSE.[3] Moreover, we note that also the Sueyoshi (1999) local criterion yields a misleading result: since $e_c = 1$, point C is denoted as CRS instead of the global sub-constant characterization. In other words, the local indicator does not detect the direction to an optimal scale size.

[2] Note that Sueyoshi's (1999) "multiple solutions" are the multiple values of the scale elasticity at a corner point (vertex) of the frontier, and not the multiple CRS solutions to the scale-efficiency problem. The relationship between this kind of multiplicity and the multiplicity of solutions in a reference set is discussed in Sueyoshi and Sekitani (2007), but only for production-based returns to scale.

[3] See the numerical example in Cesaroni et al. (2017b, p. 1444).

Fig. 1 Multiple CRS solutions (left) and single CRS solution (right) with non-monotonic RAC



This kind of divergence between local and global scale economies indicators could also take the form illustrated in Figure 1 (right). Herein, each point along the curve-section BC operates under local increasing scale economies ($e_c > 1$) while these points are actually in a DRS regime from the global point of view - given that point A is their optimal scale size.

For these reasons, it is important to integrate CRS multiple-solutions in a global classification method and to determine the form of the RAC curve to check the correspondence between local and global indicators in a convex technology. This is exactly what we do in the following sections.

3. A Ray Average Cost Model

Cesaroni and Giovannola (2015) demonstrate that a dual measure of returns to scale can be based on minimizing RAC on a strongly disposable VRS technology (either convex or non-convex). In the remainder, we propose a method for convex technologies which pays special attention to the eventual occurrence of multiple optimal solutions and to its analogy with the previously mentioned analyses. The concept of scale economies we adopt follows the standard approach and determines the behavior of the cost function in response to equi-proportional variations in outputs for given input prices (e.g., Baumol et al. 1982). Constant scale economies arise when average cost remains stationary at a minimum level. Otherwise, we have increasing or decreasing scale economies depending on whether the average cost is decreasing with an increase or a decrease in the output's scale size, respectively (see Sueyoshi 1999 (section 4.2.1), and Tone and Sahoo 2003 (section 4.2)). The exact nature of scale economies results both from the technological and organizational factors which determine the frontier of technology at different levels of outputs (see the definition by Silberton in Tone and Sahoo 2003, p. 167).

3.1 New Classification Method for Scale Economies

Evaluation of the overall efficiency (i.e., its CRS cost efficiency) and scale efficiency of the output mix of a DMU *j* can be done by means of a RAC ratio or average-cost efficiency (ACE). The ACE ratio compares the RAC evaluated at a unit $(\mathbf{x}_h, \mathbf{y}_h) \in T_{VRS}$ and that evaluated at the DMU's current scale size. This ACE ratio is defined as follows

$$R_j \equiv \frac{\mathbf{p}_j \mathbf{x}_h}{\mathbf{p}_j \mathbf{x}_j} \cdot \gamma_{j,h} \tag{3}$$

where $\gamma_{j,h}$ is the radial scaling factor comparing the output vector of DMU j with that of DMU *h*. This radial scaling factor can be computed as

$$\gamma_{j,h} = \max_r \left\{ \frac{y_{rj}}{y_{rh}} \right\}, \text{ where } \gamma_{j,h} \in (0, \infty] \tag{4}$$

For a given unit *h*, the ACE ratio R_j specifies the gain in average cost that a single DMU *j* achieves when it changes its scale size by adopting $\mathbf{x}_h$ to produce $\mathbf{y}_h$ (or equivalently, the radial projection of its output vector onto the scale size of unit *h*, $\frac{1}{\gamma_{j,h}} \cdot \mathbf{y}_j$). Note that $\mathbf{x}_h$ is not a radial projection of the input mix $\mathbf{x}_j$, since input proportions may well vary. Moreover, since $\frac{1}{\gamma_{j,h}} \cdot \mathbf{y}_j \leq \mathbf{y}_h$, the VRS assumption is not violated.

The average-cost efficiency measure (ACE) R_j^* of a given DMU j is obtained from the minimization of (3) over the least restrictive technology T_{VRS}. Furthermore, the minimizer of this optimization program is called an optimal scale size (OSS) o and has the following properties (see Cesaroni and Giovannola 2015, Propositions 3 and 7 resp.):

a) An OSS is average-cost efficient: $R_o^* = 1$;

b) The ACE measure R_j^* is equal to CRS cost efficiency and can be decomposed into the product of VRS cost efficiency and cost scale efficiency.

Otherwise stated, the optimal scale size minimizing RAC under a VRS technology coincides with the scale size that minimizes the total cost of production under the CRS assumption, i.e., R_j^* is equal to $\frac{\mathbf{p}_j \mathbf{X}\mathbf{z}^*}{\mathbf{p}_j \mathbf{x}_j}$ obtained in T_{CRS} (see Definition 1). This result indicates that the ACE measure can in principle be employed to determine the global scale economies of a given DMU. For this purpose, we just exploit the information given by coefficients R_j^* and $\gamma_{j,o}$, where evidently the last coefficient indicates the direction towards an optimal scale size.

To formulate the required classification method, the issue of the multiplicity of optimal scale sizes must first be dealt with. To this end, we present a useful lemma.

Lemma 1 The optimal scale size of a DMU j is unique up to a positive multiplicative constant.

Proof: Suppose that DMU j has two different optimal scale sizes, o and o'. Compare first o' with o. From a minimum ray-average-cost property of an optimal scale size (see Proposition 3 in Cesaroni and Giovannola 2015) we have $\frac{\mathbf{p}_j \mathbf{x}_o}{\mathbf{p}_j \mathbf{x}_{o'}} \cdot \gamma_{o',o} \geq 1$, where $\gamma_{o',o} = \max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\}$. In addition, the minimization of DMU j's ray average cost implies $\mathbf{p}_j \mathbf{x}_{o'} \cdot \gamma_{j,o'} = \mathbf{p}_j \mathbf{x}_o \cdot \gamma_{j,o}$, from which we have (a1) $\frac{\mathbf{p}_j \mathbf{x}_o}{\mathbf{p}_j \mathbf{x}_{o'}} = \frac{\gamma_{j,o'}}{\gamma_{j,o}}$. We can then express the former inequality as (a2) $\frac{\gamma_{j,o'}}{\gamma_{j,o}} \cdot \max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\} \geq 1$. Moreover, the property of being a most productive scale size implies that the RAP of *o'* with respect to *o* is equal or greater than one, therefore we have (a3) $\frac{\gamma_{j,o}}{\gamma_{j,o'}} \cdot \frac{1}{\max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\}} \geq 1$ (see Proposition 4 and expression (10) in Cesaroni and Giovannola 2015). Now, note that (a3) is the reciprocal of (a2): both inequalities can hold if and only if the equality sign applies. Selecting inequality (a2), we must then have (a4) $\frac{\gamma_{j,o'}}{\gamma_{j,o}} \cdot \max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\} = 1$. The same reasoning applies in the comparison between *o* and *o'*, so that (a5) $\frac{\gamma_{j,o}}{\gamma_{j,o'}} \cdot \max_r \left\{ \frac{y_{ro}}{y_{ro'}} \right\} = 1$ follows as well. Employing (a4), (a5) can be rewritten as $\max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\} \cdot \max_r \left\{ \frac{y_{ro}}{y_{ro'}} \right\} = 1$. Because of $\frac{1}{\max_r \left\{ \frac{y_{ro}}{y_{ro'}} \right\}} \equiv \min_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\}$, (a5) can be finally read as $\max_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\} = \min_r \left\{ \frac{y_{ro'}}{y_{ro}} \right\}$, which amounts to establishing that output vectors $\mathbf{y}_{o'}$ and $\mathbf{y}_o$ are equal up to a multiplicative constant, $\mathbf{y}_{o'} = \alpha \cdot \mathbf{y}_o$. Equality (a4) can then be used to show that $\alpha = \frac{\gamma_{j,o}}{\gamma_{j,o'}}$, while $\mathbf{x}_{o'} = \alpha \cdot \mathbf{x}_o$ because of the identity of input prices in (a1). Q.E.D.

Lemma 1 reveals that proportionality of optimal scale sizes is not only a sufficient, but also a necessary condition for the GSCSE case. Multiple solutions to the minimization of RAC can occur if and only if exact proportional replicas of the DMUs acting as an optimal scale size are present in the data set, or are rather assumed to exist on the basis of a replicability postulate (see Agrell and Tind 2001).

We can now present the following key result.

Theorem 1 In T_{VRS}, a cost-efficient DMU *j* cannot exhibit global sub-constant scale economies.

Proof: The GSCSE case, which we assume to hold, can be described as $R_j^* < 1$ and $\gamma_{j,o} > 1, \gamma_{j,o'} < 1$, where *o* and *o'* denote two distinct optimal scale sizes (see Definition 1 and expression (4)). Lemma 1 has proven that each optimal scale size is a proportional replica of the other. This implies that in T_{VRS}, by taking a suitable linear convex combination of *o* and *o'*, a production possibility $\bar{o}$ can always be taken such that $R_j^* < 1$ and $\gamma_{j,\bar{o}} = 1$. Consequently, $\mathbf{y}_{\bar{o}} \geq \mathbf{y}_j$, and R_j^* coincides with the VRS cost-efficiency score. As DMU *j* is a VRS cost-efficient unit, the above coincidence implies $R_j^* = 1$, which contradicts the assumption $R_j^* < 1$. Q.E.D.

In graphical terms we point out that Theorem 1 implies that the behavior of the VRS cost function depicted in Figure 1 (left) cannot occur, because in the interval between two optimal scale sizes (i.e., points A and B) it does not deviate from the straight line denoting the CRS cost function. Incidentally, observe that this ensures that the method by Färe and Grosskopf (1985) correctly determines the global scale economies regime in the presence of multiple optimal CRS solutions.

Because of Theorem 1, the method to classify scale economies of cost-efficient points on a least restrictive technology T_{VRS} can now be formulated as follows:

Theorem 2 For a cost-efficient DMU *j*, we have:

(i) If $R_j^* = 1$ and $\gamma_{j,o} = 1$ in a solution, then global constant scale economies prevail;

(ii) If $R_j^* < 1$ and $\gamma_{j,o} < 1$, then global increasing scale economies prevail;

(iii) If $R_j^* < 1$and $\gamma_{j,o} > 1$, then global decreasing scale economies prevail.

Proof: The possible existence of multiple solutions to the minimization of (3) implies that multiple values of $\gamma_{j,o}$ might be associated to a given R_j^*. Therefore, case (i), which includes the case[4] of $R_j^* = 1$ and $\gamma_{j,o} \neq 1$, can be immediately derived from the above properties under a) and b), while cases (ii) and (iii) follow from expression (4) and the impossibility of the simultaneous occurrence of $R_j^* < 1$ and $\gamma_{j,o} = 1$ (see Proof of Theorem 1). Q.E.D.

3.2 Discussion

Based on the properties listed under points a) and b) above, two important characteristics of this method in Theorem 2 can be pointed out. First, if DMU *j* is a cost-efficient point in T_{VRS}, then R_j^* represents its cost scale efficiency: in this sense our method is based on the measurement of cost scale efficiency (just like the method of Färe and Grosskopf 1985). Second, the equivalence of ACE and CRS cost-efficiency programs implies that $\gamma_{j,o} = \sum_{j=1}^{n} z_j^*$, where the second member denotes the CRS solution to program (2) and is - *mutatis mutandis* - equivalent to $\sum_{j=1}^{n} \lambda_j^*$, the sum of weights in the optimal solution in the Banker and Thrall (1992, p. 81) method (see also Corollary 6 in Banker 1984, p. 41).

In particular, this second characteristic is essential for the determination of $\gamma_{j,o}$, because - given the convexity of T_{VRS} - direct enumeration of (3) over the set of available observations would not necessarily ensure the detection of the sought minimum. Therefore, the CRS solution to program (2) can be immediately used to determine R_j^*, $\gamma_{j,o}$ and the point of T_{VRS} acting as the

[4]See Case 2 in Banker and Thrall (1992: p. 81).

optimal scale size of observation *j*: this point is simply obtained as a convex combination of the CRS reference set, with each weight of the combination given as $z_j^* \Big/ \sum_{j=1}^{n} z_j^*$.

We conclude that our approach offers a method for the global classification of scale economies in convex production technologies requiring the solution of two programming problems (i.e., the VRS and CRS versions of problem (2)), instead of the three problems featured in the Färe and Grosskopf (1985) method. Moreover, our method supplies quantitative information on the degree of scale economies that is unavailable in the Färe and Grosskopf (1985) approach. In fact, note that $1-R_j^*$ indicates the maximum decrease in average cost associated to a variation $\frac{1}{\gamma_{j,o}}-1$ in the scale of production. These two magnitudes suffice to completely characterize the global economies of scale of any cost-efficient DMU. This is clearly shown by the fact that, similarly to a traditional cost scale elasticity[5], an appropriate finite global measure of cost scale elasticity can be defined as:

$$\bar{e}_c = \frac{1-\gamma_{j,o}}{R_j^* - \gamma_{j,o}} \tag{5}$$

Obviously, $\bar{e}_c$ differs from e_c for both being based on finite variations in outputs and cost and involving the level of outputs associated to the global optimum - instead of the current level of outputs.

[5] We employ the basic definition as ratio of the rate of increase in output to the rate of increase in cost, see expression (32) in Sueyoshi (1999, p. 1602).

4. Local and Global Scale Economies Classifications

As shown by Podinovski (2004) for the case of the RAP function, the behavior of the RAC function of a DMU's output mix is essential for ascertaining whether local and global scale economies indicators coincide or not in a convex technology.

4.1 Ray Average Cost Function

The issue to be investigated is the behavior of the minimum RAC as a function of equi-proportional variations in $\bar{\mathbf{y}}_j$, i.e., the VRS optimal output-vector that minimizes the cost of producing the output mix $\mathbf{y}_j$, $\mathbf{p}_j\bar{\mathbf{x}}_j$. In this regard, observe that Theorem 1 is partially relevant to this purpose, because it establishes that in the interval between two optimal scale sizes the RAC function is stationary (i.e., $R_j^*(\gamma_{j,o}) = 1$). Therefore, given this result, we only need to determine the behavior outside the above-mentioned range, that is to say the case where the output level $\bar{\mathbf{y}}_j$ is expanding/contracting towards the level determined by an optimal scale size $(\mathbf{x}_o, \mathbf{y}_o) \in T_{VRS}$, $\frac{1}{\gamma_{j,o}} \cdot \bar{\mathbf{y}}_j$ with $\gamma_{j,o} \neq 1$.

To this purpose, we can express any output level $\mathbf{y}' \in \left[\bar{\mathbf{y}}_j, \frac{1}{\gamma_{j,o}} \cdot \bar{\mathbf{y}}_j\right)$ as $\mathbf{y}' \equiv (1-\alpha) \cdot \bar{\mathbf{y}}_j + \alpha \cdot (\frac{1}{\gamma_{j,o}} \bar{\mathbf{y}}_j)$, where $\alpha \in [0,1)$ is a continuous control-parameter.[6] By using expression (3) in Cesaroni and Giovannola (2015, p. 122), the RAC of any output mix $\mathbf{y}'$ can thus be expressed as

[6] For the ease of exposition, we are assuming that $\gamma_{j,o} < 1$, but the same representation obviously holds for the opposite case $\gamma_{j,o} > 1$.

$$\frac{\mathbf{p}_j \bar{\mathbf{x}}_j(\alpha)}{t} \tag{6}$$

where $t \equiv \frac{\mathbf{y}'}{\bar{\mathbf{y}}_j}$ and $\mathbf{p}_j \bar{\mathbf{x}}_j(\alpha)$ is the VRS optimal cost of producing $\mathbf{y}'$. We can now point out that the problem under discussion may be solved by analyzing the derivative of expression (6) with respect to α. This leads to the following important result.

Theorem 3 The RAC of $\mathbf{y}'$ is monotonically decreasing and convex in either an expansion or a reduction to the output level of an optimal scale size when technology is convex.

Proof: Shephard (1970) proves that the cost function is convex in the outputs if and only if the production technology is convex. This allows to get equality $\mathbf{p}_j \bar{\mathbf{x}}_j(\alpha) = (1-\alpha)\cdot \mathbf{p}_j \bar{\mathbf{x}}_j + \alpha \cdot \mathbf{p}_j \mathbf{x}_o + \delta(\alpha)$, with $0 \le \alpha < 1$ and $\delta(\alpha) \le 0$. Thus, we can express the ray average cost (6) as

$$\frac{\mathbf{p}_j \bar{\mathbf{x}}_j + \alpha \cdot (\mathbf{p}_j \mathbf{x}_o - \mathbf{p}_j \bar{\mathbf{x}}_j) + \delta(\alpha)}{1 + \alpha \cdot (\frac{1}{\gamma_{j,o}} - 1)} \tag{7}$$

The proof is made up of two parts.

Part 1: Monotonicity

By differentiating (7) with respect to α and rearranging, we obtain the final expression of the derivative as

$$\frac{(\mathbf{p}_j \mathbf{x}_o - \frac{1}{\gamma_{j,o}} \cdot \mathbf{p}_j \bar{\mathbf{x}}_j) + [\delta'(\alpha) + (\frac{1}{\gamma_{j,o}} - 1)\cdot(\alpha\delta'(\alpha) - \delta(\alpha))]}{\left[1 + \alpha \cdot (\frac{1}{\gamma_{j,o}} - 1)\right]^2} \tag{8}$$

where $\delta'(\alpha)$ denotes the derivative of δ with respect to α.

Being $\delta(0)=0$, (8) evaluated at $\alpha=0$ reduces to

$$(\mathbf{p}_j\mathbf{x}_o - \frac{1}{\gamma_{j,o}}\cdot\mathbf{p}_j\bar{\mathbf{x}}_j) + \delta^{'}(0) \tag{9}$$

This expression is negative because of: a) $(\mathbf{p}_j\mathbf{x}_o - \frac{1}{\gamma_{j,o}}\cdot\mathbf{p}_j\bar{\mathbf{x}}_j) < 0$, where this last inequality comes from the assumption of $(\mathbf{x}_o, \mathbf{y}_o)$ being the optimal scale size of $(\bar{\mathbf{x}}_j, \bar{\mathbf{y}}_j)$: $\gamma_{j,o}\cdot\mathbf{p}_j\mathbf{x}_o < \mathbf{p}_j\bar{\mathbf{x}}_j$, and from the positivity of $\gamma_{j,o}$; and b) $\delta^{'}(0)\le 0$ - resulting from the convexity of the cost function in the outputs.

The derivative of the ray average cost function is negative at $\alpha=0$ - i.e. at the initially selected value $\bar{\mathbf{y}}_j$ - but this value can be chosen to be arbitrarily close to or distant from $\frac{1}{\gamma_{j,o}}\cdot\bar{\mathbf{y}}_j$, therefore negativity holds for each $\mathbf{y}^{'}\in\left[\bar{\mathbf{y}}_j, \frac{1}{\gamma_{j,o}}\cdot\bar{\mathbf{y}}_j\right)$. Note moreover that this conclusion is independent from $\gamma_{j,o}<1$: it identically holds for $\gamma_{j,o}>1$, i.e., for each $\mathbf{y}^{'}\in\left(\frac{1}{\gamma_{j,o}}\cdot\bar{\mathbf{y}}_j, \bar{\mathbf{y}}_j\right]$.

Part 2: Convexity

The second derivative of the ray average cost with respect to α can be obtained from the differentiation of (8). After some manipulations, we get its final expression as

$$\frac{\delta^{''}(\alpha)}{\left[1+\alpha\cdot(\frac{1}{\gamma_{j,o}}-1)\right]} - \frac{2\cdot num\cdot(\frac{1}{\gamma_{j,o}}-1)}{\left[1+\alpha\cdot(\frac{1}{\gamma_{j,o}}-1)\right]^3} \tag{10}$$

where $\delta^{''}(\alpha)$ indicates the second derivative with respect to α and *num* is the numerator of (8). Now, at $\alpha=0$, (10) reduces to

$$\delta^{''}(0) - 2\cdot num(0)\cdot(\frac{1}{\gamma_{j,o}}-1) \tag{11}$$

where *num(0)* is expression (9), which has a negative sign. Moreover, note that convexity of the cost function in the outputs implies $\delta^{''}(0) \geq 0$ and $\delta^{''}(0) \leq 0$ for $\gamma_{j,o} < 1$ and $\gamma_{j,o} > 1$, respectively.

Therefore, (11) is positive for $\gamma_{j,o} < 1$, and negative for $\gamma_{j,o} > 1$. This implies that the ray average cost function is convex in both an expansion and a contraction to an optimal scale size. Q.E.D.

The following remarks can be made:

- The algebraic proof employs a Shephard (1970) result on the cost function to determine the sign of the first and second derivative of (6) with respect to α at $\alpha = 0$. As long as $\mathbf{p}_j \bar{\mathbf{x}}_j(\alpha)$ and *t* are positive - i.e., a minimum RAC exists[7] - the proof holds for any convex production technology including the case of non-parametric (polyhedral) production technologies. In graphical terms, we point out that Theorem 3 implies that the behavior depicted in Figure 1 (right) cannot occur: i.e., also the curve-section to the right of point A is monotonically increasing.

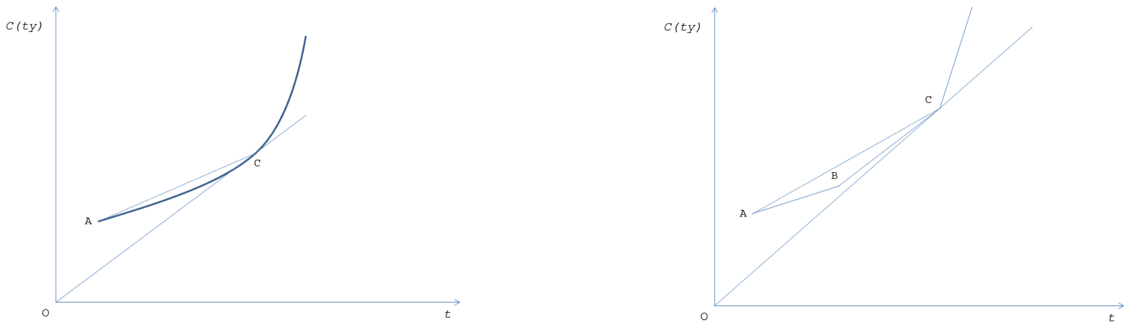
- The cost function generated by a polyhedral technology corresponds to the case $\delta^{''}(0) = 0$.

Figure 2 illustrates the algebraic sign of $\delta^{'}(0)$ and $\delta^{''}(0)$ implied by the cost function of a convex technology to graphically illustrate Theorem 3. In Figure 2 (left), point A is a generic cost-scale inefficient point, while C denotes its optimal scale size; the dashed line is the linear approximation of the cost function: $(1-\alpha) \cdot \mathbf{p}_j \bar{\mathbf{x}}_j + \alpha \cdot \mathbf{p}_j \mathbf{x}_o$. Clearly, at point A, $\delta^{'}(0) < 0$ and $\delta^{''}(0) > 0$. Differently from the smooth case, in Figure 2 (right) we have two kinds of cost-scale inefficient points, i.e., point A and point B. For point A, we note that $\delta^{'}(0) < 0$ and $\delta^{''}(0) = 0$,

[7] See Chavas and Cox (1999, p. 307): footnote 18.

while for point B we have $\delta^{'}(0)=0$ and $\delta^{''}(0)=0$, because in this last case the linear approximation of the cost function coincides with the line segment BC.

Fig. 2 Cost function of a smooth technology (left) and a polyhedral technology (right)



Observe that in Figure 2 (right), we only need to assume that the right derivatives $\delta^{'}(0)$ and $\delta^{''}(0)$ exist. Conversely in the case of a contraction to the output level of an optimal scale size (i.e., starting from a point located above C), we only need to assume that left derivatives exist. In other words, we are not imposing differentiability of $\delta(.)$ and $\delta^{'}(.)$ in the proof of Theorem 3.

We conclude this sub-section by remarking that Theorems 1 and 3 prove that in a convex production technology the RAC function is convex (u-shaped). To the best of our knowledge, this is a finding that is not yet present in the literature, as paradigmatically shown by, e.g., Baumol and Fisher (1978), Førsund and Hjalmarsson (2004) and Ray (2015) where a convex production technology is assumed, but no general conclusion on the form of the RAC function is supplied.

Remark 1 Obviously, the RAC of $\mathbf{y}^{'}$ is not monotonically decreasing in either an expansion or a reduction to the output level of an optimal scale size when technology is non-convex. This is ultimately based on the Shephard (1970, p. 227) result on the convexity of the cost function in the outputs if and only if the technology is convex. More recently, Briec, Kerstens and Vanden Eeckaut (2004) provide a more detailed result stating that for a given returns to scale assumption cost functions defined relative to convex technologies are smaller than or equal to cost functions defined relative to non-convex technologies: equality only holds in the single output and constant returns to scale case. [8] A graphical illustration of Remark 1 is provided in section 5.

4.2 Discussion

Summing up, Theorem 1 excludes the possibility of GSCSE - by showing that a point with a cost-scale elasticity equal to one is necessarily cost scale efficient -, while Theorem 3 states that a cost-scale inefficient output mix $\mathbf{y}'$ features an elasticity measure which is always greater (lower) than one in an expansion (contraction) towards its optimal scale size. Therefore, these two theorems extend the coincidence of local and global returns-to-scale classifications in convex technologies from production to cost analysis. This result has - to the best of our knowledge - not been delivered so far in the literature.

As a consequence, several remarks are in order. First, the Sueyoshi (1999) local method can be employed not only to provide local quantitative information on the degree of scale economies in a neighborhood of a DMU's output mix, but also to determine its global scale economies classification from a qualitative point of view, i.e., without providing any quantitative information on the global optimum of the VRS cost function. However, from the computational point of view, it must be stressed that this method is unduly cumbersome if used only for purposes of global classification, because of the complexity of its dual problem which - in addition - requires to be solved twice in case of an original observation which is VRS cost-efficient (see section 3.5 in Sueyoshi 1999, p. 1601).

[8] Examples of studies illustrating empirical differences between cost functions estimated relative to convex and non-convex technologies are, e.g., Balaguer-Coll, Prior and Tortosa-Ausina (2007) and De Borger and Kerstens (1996).

Second, in this last regard, both the Färe and Grosskopf (1985) and our own method are more straightforward options, with our method being more convenient as a result of both a greater computational ease and the quantitative information $\bar{e}_c$ supplied on the global optimum.

Third, our approach can provide a local measure which is a substitute for the Sueyoshi (1999) measure e_c. This can be derived from the same definition underlying (5), by considering a level of output $\mathbf{y}'$ close to $\bar{\mathbf{y}}_j$ along the direction to an optimal scale size, i.e., a predetermined and small finite distance: e.g., t=1.02 (increasing regime case) or 0.98 (decreasing regime case), etc. In particular, we obtain

$$\bar{e}_{cl} = \frac{t-1}{r_c - 1} \tag{12}$$

where $r_c \equiv \dfrac{\mathbf{p}_j \bar{\mathbf{x}}_j(t)}{\mathbf{p}_j \bar{\mathbf{x}}_j}$ is the cost ratio computed along the VRS cost function, and $0 < |t-1| < \left| \dfrac{1}{\gamma_{j,o}} - 1 \right|$. Note that the numerator of this ratio can be determined by solving a simple cost-efficiency problem for output levels $t \cdot \bar{\mathbf{y}}_j$ in T_{VRS}.

We point out that measure (12) achieves a significant advantage over the Sueyoshi (1999) measure e_c, because it refers to finite variations in outputs, rather than to infinitesimal variations which have no practical relevance in the management of production, and - as a consequence - it can also be computed for several output targets (i.e., values of t). Thus, these two features are very important for decision support by granting the firm manager a precise measure which is not biased, or limited, by the assumption of an infinitesimal variation in production.

5. Empirical Applications

5.1. Sueyoshi (1997) Data

The secondary data set presented in Sueyoshi (1997, pp. 788-789) relates to the Japanese Nippon Telegraph & Telephone company. It contains 39 time-series observations over the period April 1953 through March 1992 (fiscal years). There is a production technology with three inputs (total assets, total access lines, and total number of employees) generating three outputs (toll telephone services, local telephone services, and other miscellaneous revenues). This data set is used to empirically illustrate both the use of our proposed method as an alternative to those of Färe and Grosskopf (1985) and Sueyoshi (1999) and the coincidence between global and local indicators of scale economies under convexity.

5.1.1 Estimates of Global Scale Economies

The application of the procedure discussed in Section 3 yields the results summarized in Table 1. In this table, OE denotes overall efficiency (i.e., VRS cost efficiency of the original observations), R_j^* is the CRS cost efficiency of the projections of the original observations onto the efficient frontier, and GSE signifies the global scale economies regime of each projection - as determined by $\gamma_{j,o}$ (where I and C stand for increasing and constant scale economies, respectively).

The application of the Färe and Grosskopf (1985) method yields a global classification that is identical to the one of Table 1. This is shown by the fact that since $\gamma_{j,o}$ is always smaller or equal to unity ($\gamma_{j,o} \leq 1$), each R_j^* can also be interpreted as the NIRS cost efficiency score of the projections, whose VRS cost efficiency is - by definition - equal to one. This coincidence in global classifications clearly shows that there is no need to solve the NIRS problem, as suggested by our method: in the data set under examination, we can dispense with thirty-nine optimization problems. In addition, as can be seen from Table 1, our procedure provides the quantitative information on the global optimum given by $\bar{e}_c$ and $\gamma_{j,o}$, which the Färe and Grosskopf (1985) method does not offer. This information allows to both get an initial estimate of the cost-scale elasticity and form an opinion about the difficulties associated to the exploitation (in a VRS technology) of the efficiency gains determined by R_j^*. As an example, consider DMUs 69/70 and 82/83: both units are provided with an accurate estimate of cost scale elasticity, but while the former features an 11.8% decrease in in its average cost associated to a 400% increase in outputs, the latter exhibits a 4% decrease of the average cost brought by a 24% increase in outputs. This clearly indicates a very restrictive condition for the first DMU to implement the notional reduction in its average cost. Note that these evaluations are impossible in the Färe and Grosskopf (1985) setting, where a sole efficiency measure is made available and - moreover - at a computational cost greater than that of our method.

5.1.2 Estimates of Local Scale Economies

The method of Sueyoshi (1999) based on the cost scale elasticity e_c yields a local scale economies classification identical to the global one shown in the left section of Table 1 (cfr., ibid., Table 2 p. 1606). This empirically confirms our Theorem 3 on the coincidence, for a cost-scale inefficient output mix, of local and global indicators of scale economies in a convex production technology.

Table 1. Global and local scale economies (RAC method) for Sueyoshi (1997)

	Global scale economies				Finite global scale elasticity	Local cost scale elasticities		
DMU (year)	OE	R_j^*	$\gamma_{j,o}$	GSE	$\bar{e}_c$	$\bar{e}_{cl}$	e_c	LSE
1953/54	1.000	0.401	0.018	I[#]	2.564	2.564	2.564	I
54/55	1.000	0.379	0.021	I	2.731	3.048	3.048	I
55/56	0.977	0.389	0.023	I	2.672	2.931	2.979	I
56/57	0.972	0.446	0.028	I	2.324	2.497	2.540	I
57/58	0.958	0.466	0.032	I	2.232	2.374	2.412	I
58/59	0.937	0.472	0.036	I	2.212	2.325	2.371	I
59/60	0.933	0.503	0.043	I	2.080	2.162	2.205	I
60/61	0.906	0.524	0.050	I	2.003	2.063	2.105	I
61/62	0.885	0.553	0.060	I	1.907	1.948	1.988	I

62/63	0.833	0.569	0.066	I	1.856	1.885	1.926	I
63/64	0.814	0.604	0.078	I	1.752	1.772	1.808	I
64/65	0.796	0.634	0.092	I	1.676	1.689	1.721	I
65/66	0.769	0.665	0.105	I	1.599	1.607	1.635	I
66/67	0.774	0.708	0.126	I	1.502	1.507	1.530	I
67/68	0.776	0.735	0.151	I	1.452	1.454	1.475	I
68/69	0.776	0.764	0.174	I	1.399	1.399	1.418	I
69/70	0.781	0.782	0.203	I	1.378	1.378	1.393	I
70/71	0.765	0.793	0.230	I	1.367	1.367	1.381	I
71/72	0.749	0.804	0.260	I	1.361	1.361	1.373	I
72/73	0.735	0.817	0.302	I	1.355	1.355	1.365	I
73/74	0.737	0.835	0.352	I	1.341	1.341	1.349	I
74/75	0.713	0.833	0.382	I	1.369	1.369	1.376	I
75/76	0.718	0.852	0.429	I	1.350	1.350	1.355	I
76/77	0.770	0.887	0.509	I	1.300	1.300	1.305	I
77/78	0.896	0.932	0.658	I	1.250	1.250	1.252	I
78/79	0.888	0.938	0.690	I	1.251	1.251	1.253	I
79/80	0.889	0.945	0.727	I	1.251	1.251	1.252	I
80/81	0.881	0.948	0.749	I	1.260	1.260	1.252	I
81/82	0.878	0.952	0.766	I	1.262	1.262	1.264	I
82/83	0.889	0.959	0.803	I	1.267	1.267	1.268	I
83/84	0.907	0.966	0.839	I	1.271	1.271	1.272	I
84/85	0.905	0.967	0.852	I	1.286	1.286	1.288	I
85/86	0.935	0.980	0.912	I	1.302	1.302	1.303	I
86/87	0.965	0.990	0.954	I	1.298	1.298	1.299	I
87/88	1.000	1.000	1.000	C#	1.000	1.000	1.000	C
88/89	1.000	1.000	1.000	C	1.000	1.000	1.000	C
89/90	0.996	0.999	0.999	I	1.315	n.a.	1.315	I
90/91	1.000	1.000	1.000	C	1.000	1.000	1.000	C
91/92	1.000	1.000	1.000	C	1.000	1.000	1.000	C

I = Increasing scale economies; C = Constant scale economies; n.a.= not available.

Quite interestingly, we can now proceed to illustrate the use of our method as a local one and compare its results to those of Sueyoshi (1999). As an example, we compute our measure $\bar{e}_{cl}$ at the level of *t=1.01* for each of the cost-scale inefficient units, an exception being made for unit 89/90 for which *t=1.0003* - due to the inequality mentioned as part of (12). The results are presented in the last three columns of Table 1 headlined as “Local cost scale elasticities”, where LSE denotes the local scale economies classification.

Our local measure delivers the same classification as the Sueyoshi (1999) method. However, we can remark that the two elasticity measures generally differ as a consequence of the use of finite variations in $\bar{e}_{cl}$: the maximal difference (e_c - $\bar{e}_{cl}$) is equal to 0.071 while the average is 0.016, thus indicating an overestimation caused by Sueyoshi's measure (note that underestimation occurs only for DMU 80/81). Finally, we point out that in the data set at hand our approach avoids solving twelve optimization programs, which Sueyoshi's method requires for the computation of the intercept of the supporting hyperplane at corner observations (i.e., two programs for each of the six VRS cost-efficient observations, see eq. (30) in Sueyoshi 1999, p. 1601).

5.2. Cesaroni et al. (2017b) Data

To obtain additional evidence on the differences between these two competing elasticity-measures, we compute them on another data set used in Cesaroni et al. (2017b). It concerns a three inputs-two outputs production technology associated to the Italian local-public-transit sector. The results are summarized in Table 2 where elasticity $\bar{e}_{cl}$ is computed at levels t=1.01 and t=0.99 for increasing and decreasing global scale-economies points - respectively.

Table 2. Global and local scale economies (RAC method) for Cesaroni et al. (2017b)

	Global scale economies				Finite global scale elasticity	Local cost scale elasticities		
DMU	OE	R_j^*	$\gamma_{j,o}$	GSE	$\bar{e}_c$	$\bar{e}_{cl}$	e_c	LSE
1	0.636	0.998	0.976	I#	1.095	1.095	1.141	I
2	1.000	0.931	4.546	D#	0.981	0.981	0.981	D
3	0.564	0.992	1.157	D	0.952	0.952	0.823	D
4	0.579	0.929	0.553	I	1.190	1.190	1.286	I
5	0.718	0.794	4.616	D	0.946	0.676	0.676	D
6	0.713	0.975	0.799	I	1.143	1.143	1.201	I
7	0.629	0.994	0.940	I	1.111	1.111	1.163	I
8	0.627	0.975	1.535	D	0.955	0.955	0.848	D
9	1.000	1.000	1.000	C#	1.000	n.a.	1.000	C
10	0.605	0.858	0.366	I	1.288	1.288	1.486	I
11	0.795	0.946	2.976	D	0.973	0.973	0.910	D
12	0.664	0.861	0.347	I	1.269	1.269	1.508	I
13	0.467	0.967	1.733	D	0.957	0.957	0.854	D
14	0.635	0.915	0.480	I	1.195	1.195	1.322	I
15	0.533	0.945	3.994	D	0.982	0.982	0.934	D

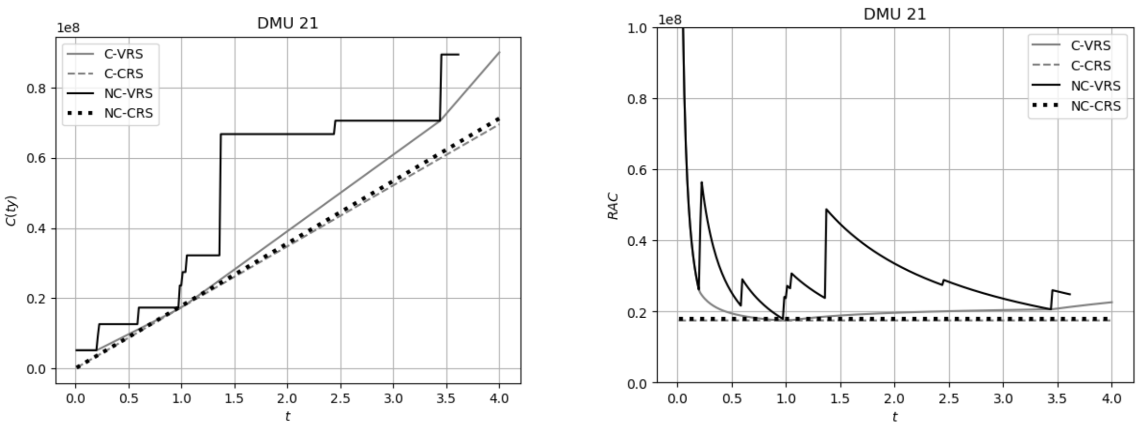
16	0.660	0.991	0.915	I	1.122	1.122	1.176	I
17	0.971	0.915	0.523	I	1.218	1.218	1.329	I
18	1.000	0.809	0.304	I	1.379	1.379	1.379	I
19	0.738	0.997	0.971	I	1.098	1.098	1.146	I
20	0.699	0.981	0.820	I	1.118	1.118	1.172	I
21	0.737	0.999	1.018	D	0.932	0.932	0.932	D
22	1.000	0.512	16.000	D	0.968	0.906	0.906	D
23	0.928	0.934	4.230	D	0.980	0.980	0.931	D
24	0.380	0.940	2.868	D	0.969	0.969	0.904	D
25	0.664	0.868	0.393	I	1.278	1.278	1.428	I
26	0.530	0.989	0.868	I	1.094	1.094	1.146	I
27	0.429	0.959	2.251	D	0.968	0.968	0.892	D
28	0.756	0.972	1.524	D	0.949	0.949	0.830	D
29	0.878	0.874	4.568	D	0.966	0.600	0.600	D
30	0.562	0.949	0.614	I	1.151	1.151	1.232	I
31	0.717	0.991	1.150	D	0.944	0.944	0.815	D
32	1.000	1.000	1.000	C	1.000	n.a.	1.000	C
33	0.545	0.936	0.549	I	1.164	1.164	1.257	I
34	1.000	0.555	17.564	D	0.974	0.922	0.922	D
35	0.795	0.956	1.866	D	0.952	0.952	0.829	D
36	0.577	0.925	0.509	I	1.181	1.181	1.267	I
37	0.736	0.989	1.165	D	0.936	0.936	0.798	D
38	0.649	0.982	0.820	I	1.109	1.109	1.162	I
39	0.875	0.601	8.174	D	0.947	0.947	0.827	D
40	0.735	0.936	3.101	D	0.971	0.971	0.910	D
41	0.641	0.979	1.617	D	0.967	0.967	0.862	D
42	0.543	0.953	4.288	D	0.986	0.986	0.942	D
43	0.818	0.952	2.452	D	0.968	0.968	0.892	D

I = Increasing scale economies; C = Constant scale economies; D = Decreasing scale economies

Differences between e_c and $\bar{e}_{cl}$ are more pronounced when compared to those shown in Table 1. First, we have a significant number of both over- and underestimation cases (21 and 19 respectively). Second, both the absolute maximal difference and the average value of the absolute differences notably increase with respect to those obtained in the Sueyoshi data set, reaching the respective values of 0.239 and 0.076. Finally, also in this data set our method obtains a computational saving of twelve optimization programs (two for each of the VRS cost-efficient observation in Table 2).

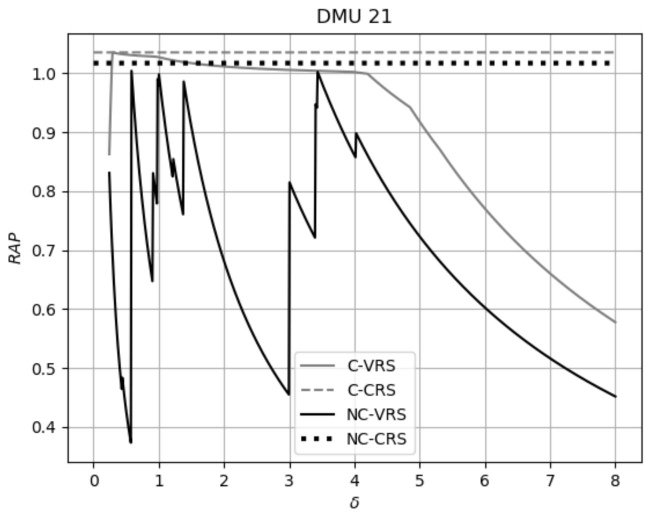
Figure 3 illustrates the above Remark 1 with reference to the output mix of DMU 21 taken from the Cesaroni et al. (2017b) data set by comparing the VRS and CRS cost functions under the convex (C) and the non-convex (NC) technologies (the latter can be obtained by adding $z_j \in \{0,1\}$ to (1)). Figure 3 (left) plots $C(ty)$ against t (just like in Figures 1 and 2), while Figure 3 (right) plots RAC against t. Starting from positive values of t near to 0 and expanding the output mix towards 1, the convex VRS (C-VRS) technology clearly exhibits a monotonically decreasing RAC (i.e., the ratio $C(ty)/t$ is convex). By contrast, in addition to the discontinuities generated by the horizontal segments of its cost function, the non-convex VRS (NC-VRS) technology presents a non-monotonic behavior of its RAC, which is increasing/decreasing both for t<1 and t>1 (i.e., the ratio $C(ty)/t$ is nonconvex). Thus, not only cost functions behave differently depending on the convexity or non-convexity of the underlying technology, but also the RAC function shares this same property.

Fig. 3 Cost function C(ty) (left) and RAC (right) for DMU 21



Finally, Figure 4 depicts the RAP of the output mix of DMU 21. Broadly speaking, the shape of RAP is the inverse of RAC. However, the comparison between the inverse of Figure 3 (left) and Figure 4 reveals that for DMU 21, both in the convex and non-convex cases, the two functions do not coincide. This is a consequence of the potential presence of allocative inefficiency in RAC, an element which does not enter the definition of RAP (see Cesaroni and Giovannola, 2015).

Fig. 4 RAP of DMU 21



6. Conclusions

This contribution advances the literature on convex non-parametric production technologies in several ways. First, it determines the u-shape of the RAC function and extends the coincidence of local and global returns to scale from production to cost analysis. Second, it introduces a global method for the determination of scale economies which simplifies the Färe and Grosskopf (1985) procedure and supplies quantitative information that is otherwise not available. Third, it proposes a local criterion for the determination of scale economies which is computationally simpler than the Sueyoshi (1999) method, while offering quantitative information which is more relevant than the cost scale elasticity employed in the latter. An empirical illustration documents these results.

References

Agrell, P.J., J. Tind (2001) A Dual Approach to Nonconvex Frontier Models, *Journal of Productivity Analysis 16*, 129-147.

Balaguer-Coll, M.T., Prior, D., Tortosa-Ausina, E. (2007) On the determinants of local government performance: A two-stage nonparametric approach, *European Economic Review* 51, 425-451.

Banker, R.D. (1984) Estimating most productive scale size using Data Envelopment Analysis, *European Journal of Operational Research 17*, 35-44.

Banker, R.D., Chang, H., Cooper, W.W. (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., Charnes A., Cooper, W.W. (1984) Models for the estimation of technical and scale inefficiencies in Data Envelopment Analysis, *Management Science 30*, 1078-1092.

Banker, R.D., Thrall, R.M. (1992) Estimation of returns to scale using Data Envelopment Analysis, *European Journal of Operational Research 62*, 74-84.

Baumol W.J., Fisher D. (1978) Cost-minimizing number of firms and determination of industry structure, *Quarterly Journal of Economics* 92, 439-468.

Baumol, W.J., Panzar, J.C., Willig, R.D. (1982) *Contestable markets and the theory of industry structure*. New York: Harcourt Brace Jovanovich.

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, 155-192.

Chavas, J.P., Cox, T.L. (1999) A generalized distance function and the analysis of production efficiency, *Southern Economic Journal 66*, 294-318.

Cesaroni, G., Giovannola, D. (2015) Average-cost efficiency and optimal scale sizes in non-parametric analysis, *European Journal of Operational Research 242*, 121-133.

Cesaroni, G., Kerstens, K., Van De Woestyne, I. (2017a) Global and local scale characteristics in convex and nonconvex nonparametric technologies: A first empirical exploration, *European Journal of Operational Research 259*, 576-586.

Cesaroni, G., Kerstens, K., Van De Woestyne, I. (2017b) Estimating scale economies in non-convex production models, *Journal of the Operational Research Society 68*, 1442-1451.

De Borger, B., Kerstens, K. (1996) Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches, *Regional Science and Urban Economics* 26, 145-170.

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

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

Førsund, F., Hjalmarsson, L. (2004) Are all scales optimal in DEA? Theory and empirical evidence, *Journal of Productivity Analysis 21*, 25-48.

Førsund, F., Kittelsen S., Krivonozhko V. (2009) Farrell revisited-Visualizing properties of DEA production frontiers, *Journal of the Operational Research Society 60*, 1535-1545.

Panzar, W., Willig, R.D. (1977) Economies of scale in multi-output production, *Quarterly Journal of Economics 91*, 481-493.

Podinovski, V. (2004) Efficiency and global scale characteristics on the “No free lunch” assumption only, *Journal of Productivity Analysis 22*, 227-257.

Ray, S. (2015) Nonparametric measures of scale economies and capacity utilization: An application to U.S. manufacturing, *European Journal of Operational Research 245*, 602-611.

Shephard, RW (1970) *Theory of cost and production functions.* Princeton, New Jersey: Princeton University Press; 1970.

Sueyoshi, T. (1997) Measuring efficiencies and returns to scale of Nippon Telegraph and Telephone in production and cost analyses, *Management Science* 43, 779-796.

Sueyoshi, T. (1999) DEA duality on returns to scale in production and cost analyses: An occurrence of multiple solutions and differences between production-based and cost-based RTS estimates, *Management Science 45*, 1593-1608.

Sueyoshi, T., Sekitani K. (2007) The measurement of returns to scale under a simultaneous occurrence of multiple solutions in a reference set and a supporting hyperplane, *European Journal of Operational Research 181*, 549-570.

Tone, K., Sahoo, B.K. (2003) Scale indivisibilities and production function in Data Envelopment Analysis, *International Journal of Production Economics 84*, 165-192.