Applied Economics, 2012, **44**, 4125–4141

Dataset is available as supplementary content

Static efficiency decompositions and capacity utilization: integrating economic and technical capacity notions

Bruno De Borger[a], Kristiaan Kerstens[b,*], Diego Prior[b,c] and Ignace Van de Woestyne[d]

[a]*Department of Economics, University of Antwerp, Prinsstraat 13, Antwerp B-2000, Belgium*
[b]*IESEG School of Management, CNRS-LEM (UMR 8179), 3 rue de la Digue, Lille F-59000, France*
[c]*Department of d'Economia de l'Empresa, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*
[d]*Faculty of Economics and Management, Hogeschool-Universiteit Brussels, Stormstraat 2, B-1000 Brussels, Belgium*

Starting from existing static decompositions of overall economic efficiency on nonparametric production and cost frontiers, this article proposes more comprehensive decompositions including several cost-based notions of capacity utilization. Furthermore, in case prices are lacking, we develop additional decompositions of overall technical efficiency integrating a technical concept of capacity utilization. These new efficiency decompositions provide a link between short and long run economic analysis and, in empirical work, avoid conflating inefficiency and differences in capacity utilization. An empirical analysis using a monthly panel of Chilean hydroelectric power plants illustrates the potential of these decomposition proposals.

Keywords: efficiency; capacity utilization

JEL Classification: D24; M21

I. Introduction

The analysis of efficiency and productivity based on frontier specifications of technology has become a standard empirical tool serving academic, regulatory and managerial purposes. Apart from its widespread application in analysing private and public sector performance-related issues (e.g. Balcombe *et al*., 2008; Glass *et al*., 2009, among others), the implementation of incentive regulatory mechanisms (e.g. price cap regulation) using frontier-based performance benchmarks is, for instance, rather widespread in countries having liberalized their network industries (see, e.g. the survey in Jamasb and Pollitt (2001) for its use in the electricity sector). As an example of a managerial application, one can point to the Commercial Bank of

*Corresponding author. E-mail: k.kerstens@ieseg.fr

http://www.informaworld.com
DOI: 10.1080/00036846.2011.587783

Greece (CBG) which has instituted performance measurement systems for bank branches using Data Envelopment Analysis (DEA) since 1988 (Athanassopoulos and Giokas, 2000).

However, it is a bit surprising that many applications have – often implicitly – taken a long run perspective: it is assumed that all inputs and/or outputs are under managerial control. Though the possibility of focusing on a sub-vector of, for instance, inputs has been recognized for long, the frontier literature has almost completely ignored the notion of capacity utilization. As a consequence, part of what may be attributed to inefficiency, may in fact be due to the short run fixity of certain inputs.

Caves (2007) recently shows how various efficiency concepts as well as the capacity notion have contributed, among others, to a rich body of empirical knowledge on firm behaviour that is often associated with the so-called old industrial organisation literature. Indeed, there is a long tradition of empirical research on organisations focusing on capacity utilization. For instance, Ghemawat and Nalebuff (1985) show how firms' survival probability depends on the ability to adjust capacity to control production costs when demand changes. Being largely a technical datum, capacity utilization becomes an organisational factor. For example, Bonin *et al*. (1993) report that cooperative firms are able to maintain more stable production plans than noncooperative firms, which is a factor that seems related with the advantages of having stable contracts with regular partners.

This article concentrates on the development of efficiency decompositions integrating capacity utilization using nonparametric frontier technologies.[1] In this nonparametric approach, piecewise linear frontiers envelop the observations as tightly as possible subject to certain minimal production axioms.[2] More specifically, this article makes, to the best of our knowledge, two contributions. First, this is the first proposal in the literature integrating different notions of capacity utilization into a taxonomy of static efficiency concepts for nonparametric technologies.[3] Second, we integrate both primal and dual concepts of capacity into this literature on multiple output nonparametric frontiers.[4] In brief, the purpose of our contribution is to carefully disentangle between capacity utilization and various efficiency concepts in a non-parametric frontier framework that allows for a coherent treatment of both primal and dual capacity notions. Already Fuss *et al*. (1978, p. 223) stressed that fundamental production axioms are of a qualitative and nonparametric nature and therefore should ideally be tested using nonparametric technologies.[5]

This article is structured as follows. Section II summarizes the traditional static decomposition of overall economic efficiency and some less known useful extensions. The next section (Section III) introduces both economic (cost-based) and technical concepts of capacity utilization. Section IV extends the traditional efficiency decomposition by integrating this variety of capacity utilization measures. These new decompositions are related to one another, with a focus on the relations between short and long run scale efficiency and capacity utilization. In addition, decompositions of overall technical efficiency integrating a technical concept of capacity utilization are developed. The latter are particularly useful when prices are lacking. An empirical section illustrates these new decompositions for a monthly panel of Chilean hydro-electric power plants observed over a single year. Conclusions are drawn in Section VI.

II. Definition of the Static Efficiency Decomposition

Microeconomic foundations of production, cost and efficiency

To clear the ground, we start by defining technology and some basic notation. Production technology is defined by the production possibility set: $S = \{(x, y) | x$ can produce $y\}$. The input set associated with S

[1] By contrast, parametric frontiers utilize parametric, locally if not even globally flexible specifications with a finite number of parameters to estimate the underlying technology.
[2] In line with tradition, we maintain convexity throughout the analysis. Notice that Tone and Sahoo (2003) stress indivisibilities in selecting among technological options and plea in favour of using nonconvex nonparametric technologies. The latter are systematically developed in Briec *et al*. (2004). Note that in principle one can dispense with convexity in the analysis developed in this contribution.
[3] In the parametric literature various productivity decompositions have been suggested to include measures of capacity utilization (see Hulten (1986), among others). Some productivity decompositions have been recently proposed in the nonparametric frontier literature (see below).
[4] Mainly dual multiple output concepts are known in the parametric literature (e.g. Squires, 1987), while primal capacity notions are difficult to estimate. Färe (1984) shows that a primal capacity notion cannot be obtained for certain popular parametric specifications of technology (e.g. the Cobb–Douglas).
[5] This does not preclude an eventual extension of our proposals into a parametric framework.

denotes all input vectors $x \in \mathbb{R}^n_+$ capable of producing a given output vector $y \in \mathbb{R}^m_+$: $L(y) = \{x|(x, y) \in S\}$. It is often useful to partition the input vector into a fixed and variable part $(x = (x^v, x^f))$ and to make the same distinction regarding the input price vector $(w = (w^v, w^f))$.

The input set $L(y)$ associated with S satisfies some combination of the following standard assumptions (see, e.g. Färe *et al.* (1985)):

- $L1$ $\forall y \geq 0$ with $y \neq 0$, $0 \notin L(y)$ and $L(0) = \mathbb{R}^m_+$
- $L2$ Let $\{y_n\}_{n \in \mathbb{N}}$ be a sequence such that $\lim_{n \to \infty} \|y_n\| = \infty$, then $\cap_{n \in \mathbb{N}} L(y_n) = 0$
- $L3$ $L(y)$ is closed $\forall y \in \mathbb{R}^m_+$
- $L4$ $L(y)$ is a convex set $\forall y \in \mathbb{R}^m_+$
- $L5$ If $x \in L(y)$, then $\lambda x \in L(y)$, $\forall \lambda \geq 1$
- $L6$ $\forall x \in L(y)$, $u \geq x \Rightarrow u \in L(y)$
- $L7$ $L(\lambda y) = \lambda\ L(y)$, $\forall \lambda \geq 0$

Apart from the traditional regularity conditions (i.e. no free lunch and the possibility of inaction ($L1$), the boundedness ($L2$), closedness ($L3$) and convexity ($L4$) of the input set, there are three other assumptions that are often invoked. Assumption ($L5$) postulates ray (or weak) disposability of the inputs, while axiom ($L6$) imposes the more traditional assumption of strong (or free) disposal of inputs. Finally, axiom ($L7$) presents the special case of a homogenous or constant returns to scale input correspondence. Note that not all of these axioms are independent of one another: e.g. assumption ($L5$) can be deduced from axiom ($L6$).

Since we only treat the static decomposition in the input orientation, we first define the input distance function that offers a complete characterization of technology. In particular, it characterizes the input set $L(y)$ as follows:

$$D_i(x, y) = \max\{\lambda : \lambda \geq 0, x/\lambda \in L(y)\} \qquad (1)$$

We next define the radial input efficiency measure as

$$DF_i(x, y) = \min\{\lambda | \lambda \geq 0, (\lambda x) \in L(y)\} \qquad (2)$$

This measure is simply the inverse of the input distance function $(DF_i(x, y) = [D_i(x, y)]^{-1})$. Its most important properties are: (i) $0 < DF_i(x, y) \leq 1$, with efficient production on the boundary (isoquant) of $L(y)$ represented by unity; (ii) it has a cost interpretation (see Färe *et al.* (1994) for details).[6]

The cost function, a dual representation of technology, indicates the minimum expenditures to produce output vector y given a vector of semi-positive input prices $w \in \mathbb{R}^n_+$

$$C(y, w) = \min\{wx | x \in L(y)\} \qquad (3)$$

This cost function can also be written in terms of the input distance function.

This dual relation establishes the foundations for efficiency measurement.[7] Discussing a few points in more detail, it is clear that for each element of the input set $(x \in L\{y\})$ the following inequality holds:

$$C(y, w) \leq w \cdot \left(\frac{x}{D_i(x, y)}\right) \qquad (4)$$

Thus, minimal costs are smaller or equal to observed cost at the isoquant of the input set (i.e. after eliminating possible technical inefficiency). This inequality can be rewritten to obtain Mahler's inequality as follows:

$$C(y, w) \cdot D_i(x, y) \leq w \cdot x \qquad (5)$$

The transformation of this inequality into equality by introducing an allocative efficiency component $AE_i(w, x, y)$ forms the theoretical foundation for the multiplicative Farrell (1957) decomposition for measuring input efficiency

$$\frac{C(y, w)}{w \cdot x} = \frac{1}{D_i(x, y)} \cdot AE_i(x, y, w) \qquad (6)$$

The first ratio of minimal to observed costs $C(y, w)/w \cdot x$ defines a cost efficiency component (labelled overall efficiency component below). The second ratio $1/D_i(x, y)$ coincides simply to the radial measure of input technical efficiency $(DF_i(x, y))$. Finally, the component $AE_i\ (w, x, y) = C(y, w)/w \cdot x \cdot D_i(x, y)$ indicates allocative efficiency, defined in a residual way.

[6] For convenience, we stick to the traditional radial input efficiency measure. Recently, more general distance functions have been introduced to measure profit efficiency (Chambers *et al.*, 1998). Apart from the fact that these new measures lead to additive rather than multiplicative decompositions, these are related to the traditional radial efficiency measures employed here.

[7] The duality relation between input distance function and cost function is

$$C(y, w) = \min_x\{wx : D_i(x, y) \geq 1\} w > 0 \text{ and}$$
$$D_i(x, y) = \min_w\{wx : C(y, w) \geq 1\} \quad x \in L(y)$$

While $C(y, w)$ can be obtained from $D_i(x, y)$ by optimizing with respect to input quantities, $D_i(x, y)$ can be resolved from $C(y, w)$ by minimizing with respect to input prices.

Extended static efficiency decompositions in the literature

While Farrell (1957) provided the first measurement scheme for the evaluation of technical and allocative efficiency in a frontier context, Färe *et al.* (1983, 1985 pp. 3–5) offer an extended efficiency taxonomy.[8] Since technologies vary in terms of underlying assumptions (Färe *et al.*, 1994), it is useful to condition the above notation of the efficiency measure on two main assumptions: (i) the difference between constant (CRS) and variable (VRS) returns to scale technologies (convention: $C = CRS$, $V = VRS$); (ii) the distinction between strong (SD) and weak (WD) disposability assumptions (convention: $S = SD$; $W = WD$). As these proposals have become a standard way to decompose efficiency in competitive markets (see, e.g. Ganley and Cubbin, 1992), we first present the definition of their taxonomy and the ensuing operational measurement procedures.[9] Note that this and all other extended static efficiency decompositions discussed below start from the basic multiplicative decomposition (Equation 6) by varying the key assumptions on technology listed above, while respecting the basic duality relations.

Definition 1: Under the above assumptions on the input set $L(y)$, the following input-oriented efficiency notions can be distinguished.

(1) Technical Efficiency is the quantity: $TE_i(x, y) = DF_i(x, y|V, W)$.
(2) Structural Efficiency is the quantity: $STE_i(x, y) = DF_i\ (x, y|V, S)/DF_i(x, y|V, W)$.
(3) Scale Efficiency is the quantity: $SCE_i(x, y) = DF_i\ (x, y|C, S)/DF_i(x, y|V, S)$.
(4) Overall Technical Efficiency is the quantity: $OTE_i(x, y) = DF_i\ (x, y|C, S)$.
(5) Overall Efficiency is the quantity: $OE_i(x, y, w|C) = C(y, w|C)/wx$.
(6) Allocative Efficiency is the quantity: $AE_i\ (x, y, w|C) = OE_i\ (x, y, w|C)/OTE_i\ (x, y)$.

We first comment on the technological part of this efficiency taxonomy. First, technical efficiency ($TE_i(x, y)$) demands that production occurs on the boundary of technology. A producer is technically inefficient otherwise. $TE_i(x, y)$ is traditionally evaluated relative to a VRS technology with WD using $DF_i(x, y|V, W)$. Second, structural efficiency ($STE_i(x, y)$) implies that production occurs in an uncongested or 'economic' production region. Otherwise, a producer is structurally inefficient. $STE_i(x, y)$ is a derivative result of computing input efficiency relative to both SD and WD technologies imposing VRS. Third, scale efficiency ($SCE_i(x, y)$) implies that the choice of inputs and outputs is compatible with the long run ideal of a CRS technology. A producer is scale inefficient otherwise. $SCE_i(x, y)$ results from comparing an observation to CRS and VRS technologies with SD.[10] Finally, overall technical efficiency ($OTE_i(x, y)$) is the result of all three previous definitions: a producer is overall technically efficient if production occurs on the boundary of a congestion-free CRS technology; it is overall technically inefficient otherwise.

As to the value function part of the efficiency decomposition, overall efficiency ($OE_i\ (x, y, w|C)$) requires computing a long run total cost function relative to a CRS technology with SD and taking a ratio of this minimal costs to actual costs. This long run total cost function is defined as follows: $C(y, w|C) = \min\{wx|x \in L(y|C, S)\}$, and can be solved by a simple linear program. $OE_i(x, y, w|C)$ can be seen as the multiplicative result of $OTE_i(x, y)$ and allocative efficiency ($AE_i\ (x, y, w|C)$), defined as a residual term making up the gap between $OE_i(x, y, w|C)$ and $OTE_i(x, y)$. $AE_i(x, y, w|C)$ requires that there is no divergence between observed and optimal costs, revenue, profits or whatever objective the producer is assumed to pursue. Otherwise, a producer is allocatively inefficient.

Note that $OE_i(x, y, w|C)$ and $AE_i(x, y, w|C)$ imply price-dependent characterizations of efficiency, while $OTE_i(x, y)$ and its components are entirely price-independent notions. Though the underlying radial efficiency measures and cost functions are evaluated on various technologies, all these components are smaller or equal to unity. These static efficiency

[8] Other classifications include Banker *et al.* (1984) and Førsund and Hjalmarsson (1974, 1979).
[9] To simplify matters, we ignore efficiency analysis in noncompetitive settings, leading to price inefficiencies in addition to inefficiencies in quantities (e.g. Färe *et al.*, 1994; Grifell-Tatjé and Lovell, 2000; Kallio and Kallio, 2002).
[10] In addition, one can obtain qualitative information on scale economies by identifying local returns to scale. When $SCE_i(x, y) = 1$, then the unit is compatible with CRS. When $SCE_i(x, y) < 1$, then the unit does not operate with optimal size. But, one cannot know whether it is subject to increasing (IRS) or decreasing (DRS) returns to scale. By computing input efficiency also relative to a SD technology with nonincreasing returns to scale ($DF_i(x, y|N, S)$) and by exploiting the nestedness of technologies, one discriminates between IRS and DRS (Färe *et al.*, 1983): (i) IRS holds when $DF_i(x, y|C, S) = DF_i(x, y|N, S) \leq DF_i(x, y|V, S) \leq 1$; (ii) DRS holds when $DF_i(x, y|C, S) \leq DF_i(x, y|N, S) = DF_i(x, y|V, S) \leq 1$.

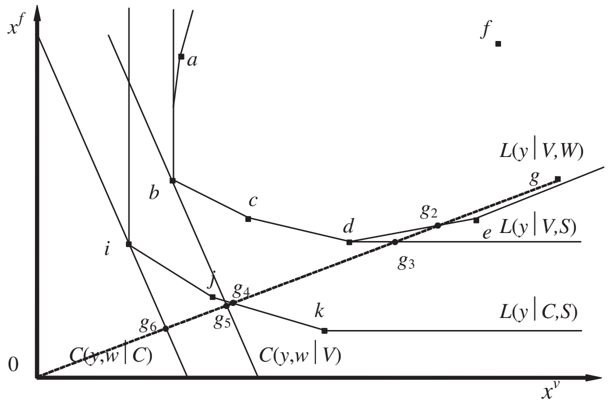

Fig. 1. DEC1 and DEC2 illustrated on input sets with different production axioms

concepts are mutually exclusive and their radial measurement yields a multiplicative decomposition:

$$OE_i(x,y,w|C) = AE_i(x,y,w|C) \cdot OTE_i(x,y) \quad \text{(DEC1)}$$

where $OTE_i(x,y) = TE_i(x,y) \cdot STE_i(x,y) \cdot SCE_i(x,y)$ (Färe *et al.*, 1985).

This traditional static efficiency decomposition is illustrated in Fig. 1 representing three technologies: one imposing SD and CRS ($L(y|C,S)$); one with SD and VRS ($L(y|V,S)$); and one with WD and VRS ($L(y|V,W)$). For observation g, $OE_i(x,y,w|C)$ is the ratio $0g_6/0g$. Its component measures are: $TE_i(x,y) = 0g_2/0g$, $STE_i(x,y) = 0g_3/0g_2$, $SCE_i(x,y) = 0g_4/0g_3$ and $AE_i(x,y,w|C) = 0g_6/0g_4$.

An alternative decomposition, proposed by Seitz (1970, 1971) but little used in practice, takes the same overall efficiency measure, and focuses on slightly different effects. It prepares the ground for the extended decompositions proposed in Section IV, since scale efficiency is based on a dual characterization of technology. His insight is that the same initial overall efficiency measure can be decomposed into several other overall efficiency measures defined with respect to different technologies. Seitz (1970, 1971) defines scale efficiency based on a dual cost function as follows.

Definition 2: Cost-based scale efficiency is defined as the quantity:

$$CSCE_i(x,y,w) = \frac{C(y,w|C)/wx}{C(y,w|V)/wx} = \frac{OE_i(x,y,w|C)}{OE_i(x,y,w|V)}$$

$CSCE_i(x,y,w)$ is a price-dependent scale efficiency term based on cost function estimates. Since $OE_i(x,y,w|C) \leq OE_i(x,y,w|V)$, $CSCE_i(x,y,w) \leq 1$.[11]

Rephrasing his proposal in the current notation, he decomposes overall efficiency ('economic efficiency' in his words) as follows:

$$OE_i(x,y,w|C) = CSCE_i(x,y,w) \cdot OE_i(x,y,w|V) \quad \text{(DEC2)}$$

where $OE_i(x,y,w|V) = TE_i(x,y) \cdot STE_i(x,y) \cdot AE_i(x,y,w|V)$. $CSCE_i(x,y,w)$ is labelled 'economic scale efficiency' by Seitz (1970, p. 508), while $OE_i(x,y,w|V)$ is termed 'economic efficiency given the scale' of operations.[12] The main difference with (DEC1) is that allocative efficiency is now defined as closing the gap between a cost function and a technical efficiency measure defined relative to a VRS instead of a CRS technology.[13] This alternative decomposition is also illustrated in Fig. 1. For observation g, $OE_i(x,y,w|C)$ is again the ratio $0g_6/0g$. Its component measures in common with (DEC1) are: $TE_i(x,y) = 0g_2/0g$ and $STE_i(x,y) = 0g_3/0g_2$. Furthermore, we now have $AE_i(x,y,w|V) = 0g_5/0g_3$ and $CSCE_i(x,y,w) = 0g_6/0g_5$.

The use of different overall efficiency measures has the advantage that each of them can be decomposed into technical and allocative efficiency components. This makes it, for instance, straightforward to link primal and dual approaches to scale efficiency. Decomposing $CSCE_t(x,y,w)$ into its technical and allocative components

$$\begin{aligned} CSCE_i(x,y,w) &= \left[\frac{DF_i(x,y|C,S)}{DF_i(x,y|V,S)}\right] \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)}\right] \\ &= SCE_i(x,y) \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)}\right] \end{aligned} \quad (7)$$

[11] Identification of local economies of scale proceeds as follows. When $CSCE_i(x,y,w) = 1$, then the unit minimizes costs and enjoys CRS. When $CSCE_i(x,y,w) < 1$, then computing a cost function relative to a nonincreasing returns to scale technology ($OE_i(x,y,w|N)$) and knowing that $OE_i(x,y,w|C) \leq OE_i(x,y,w|N) \leq OE_i(x,y,w|V) \leq 1$ (Grosskopf, 1986), the same reasoning as above applies to infer local economies of scale. This procedure applies to any dual formulation.
[12] Just as price-dependent parametric approaches have been popular in the literature, this very similar cost-based scale term has repeatedly appeared in the literature since Seitz (1970). See, for instance, Fukuyama and Weber (1999), Rowland *et al.* (1998) or Sueyoshi (1999).
[13] One could introduce the notation $AE_i(x,y,w|C)$ in (DEC1) to distinguish this component from the one in (DEC2).

Färe *et al.* (1994) show: $CSCE_i(x,y,w) = SCE_i(x,y) \Leftrightarrow AE_i(x,y,w|C) = AE_i(x,y,w|V)$[14]. Since $OE_i(x,y,w|C) \leq OE_i(x,y,w|V) \leq 1$, $[AE_i(x,y,w|C)/AE_i(x,y,w|V)] \lesseqqgtr = 1$. Furthermore, since $CSCE_i(x,y,w) \leq 1$ and $SCE_i(x,y) \leq 1$, $CSCE_i(x,y,w) \lesseqqgtr SCE_i(x,y)$.

The link between the traditional decomposition (DEC1) and the Seitz (1970, 1971) proposal (DEC2) is now easily established:

$$OE_i(x,y,w|C) = SCE_i(x,y) \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)}\right] \cdot TE_i(x,y) \cdot STE_i(x,y) \cdot AE_i(x,y,w|V) \quad (8)$$

where $OE_i(x,y,w|V)$ contains the last three terms and eliminating the common $AE_i(x,y,w|V)$ term yields (DEC 1).

Though Färe *et al.* (1985) mention a time perspective when defining scale efficiency, they mainly distinguish between private and social goals when discussing their decomposition components providing the benchmarking ideals.[15] But, an alternative interpretation is that the time perspective of organisational decision making dictates the order in which the static decomposition is defined and measured. It is important to distinguish between short and long run ideals when directing efforts for improvement. $TE_i(x,y)$ and $STE_i(x,y)$ are deemed to be short run ideals, since these goals mainly involve eliminating managerial inefficiencies. $AE_i(x,y,w)$ and especially $SCE_i(x,y)$ are long run goals: they require changes in the input mix respectively scale adjustments.

Extended static efficiency decompositions in the short-run

Since the main focus of this contribution is on establishing a link between existing efficiency decompositions and traditional capacity concepts and since capacity utilization is linked to the short term fixity of some of the inputs, it is necessary to develop a notation for efficiency measurement focusing on a sub-vector of inputs. For instance, an input efficiency measure seeking reductions in variable inputs only is defined as

$$DF_i^{SR}(x,y) = \min\{\lambda | \lambda \geq 0, (\lambda x^v, x^f) \in L(y)\} \quad (9)$$

Replicating the analysis in the previous subsection for the short run case (see, e.g. Färe *et al.* (1994, section 10.1)), one can straightforwardly develop an analogous sub-vector efficiency decomposition.

Definition 3: Under the above assumptions on the input set $L(y)$, the following short run (SR) input-oriented efficiency notions can be distinguished:

(1) SR Technical Efficiency is the quantity: $TE_i^{SR}(x,y) = DF_i^{SR}(x,y|V,W)$
(2) SR Structural Efficiency is the quantity: $STE_i^{SR}(x,y) = DF_i^{SR}(x,y|V,S)/DF_i^{SR}(x,y|V,W)$
(3) SR Scale Efficiency is the quantity: $SCE_i^{SR}(x,y) = DF_i^{SR}(x,y|C,S)/DF_i^{SR}(x,y|V,S)$
(4) SR Overall Technical Efficiency is the quantity: $OTE_i^{SR}(x,y) = DF_i^{SR}(x,y|C,S)$
(5) SR Overall Efficiency is the quantity: $OE_i^{SR}(x,y,w|C) = VC(y,w^v,x^f|C)/(w^v x^v)$
(6) SR Allocative Efficiency is the quantity: $AE_i^{SR}(x,y,w|C) = OE_i^{SR}(x,y,w|C)/OTE_i^{SR}(x,y)$

Note that the variable cost function relative to a CRS technology ($VC(y,w^v,x^f|C)$) is defined as follows: $VC(y,w^v,x^f|C) = \min\{w^v x^v | (x^v,x^f) \in L(y|C,S)\}$. It can be solved by a simple linear program. Otherwise, all comments on both the technological and value function parts of the efficiency taxonomy in the previous section carry over to these short run components. Again these static efficiency concepts taken together constitute a multiplicative decomposition

$$OE_i^{SR}(x,y,w|C) = AE_i^{SR}(x,y,w|C) \cdot OTE_i^{SR}(x,y) \quad \text{(SRDEC1)}$$

where $OTE_i^{SR}(x,y) = TE_i^{SR}(x,y) \cdot STE_i^{SR}(x,y) \cdot SCE_i^{SR}(x,y)$ and the interpretation is completely similar to (DEC 1).

For the short run case, the alternative decomposition of Seitz (1970, 1971) can now be developed as in the following definition.

Definition 4: Cost-based SR scale efficiency is defined as the quantity

$$CSCE_i^{SR}(x,y,w) = \frac{VC(y,w^v,x^f|C)/w^v x^v}{VC(y,w^v,x^f|V)/w^v x^v} = \frac{OE_i^{SR}(x,y,w|C)}{OE_i^{SR}(x,y,w|V)}$$

[14] See also Sueyoshi (1999). Actually, scale efficiency in Färe *et al.* (1994, pp. 84–7) is defined on technologies based on limited data, i.e. using information on cost data and the output vector solely. They show that scale efficiency under cost and production approaches is identical when: (i) all organizations face identical input prices; and (ii) $AE_i(x,y,w|C) = AE_i(x,y,w|V)$. When input price information is available and cost function estimates are employed, however, the first of these conditions is redundant.

[15] See Färe and Grosskopf (2000): in defense to McDonald (1996) who proposes an alternative order of some components, they justify their position by referring to economic tradition, but without mentioning a time perspective.

Again the comment in the previous section carries over to this cost-based short run scale efficiency component. The alternative short run decomposition of overall efficiency then reads:

$$OE_i^{SR}(x, y, w|C) = CSCE_i^{\mathrm{SR}}(x, y, w) \cdot OE_i^{SR}(x, y, w|V) \quad \text{(SRDEC2)}$$

where $OE_i^{SR}(x, y, w|V) = TE_i^{SR}(x, y) \cdot STE_i^{SR}(x, y) \cdot AE_i^{SR}(x, y, w|V)$ and the interpretation is again similar to (DEC2).

Closing observations

One could object that the whole decomposition is to some extent artificial in that production decisions are, at least theoretically, assumed to be taken jointly. For instance, assuming cost minimization as a realistic goal, one would expect organizations to minimize costs, and not first to decide on a technically efficient input combination and next on a technically efficient input combination that also happens to be allocative efficient.[16] But, the whole point of retrospectively benchmarking organizational performance is that organizations make judgmental errors. The decomposition then serves as a conceptual tool identifying potential sources of inefficiencies and to select realistic benchmarks to guide the improvement process. Ideally, decompositions are just identities that should be judged by their ability to guide practitioner's path to improved performance. In this perspective, capacity utilization can provide a link between the short and long run analysis. Of course, this requires a careful interpretation of the traditional capacity notions in a frontier context. We embark on this essay in economic semantics in the next section.

Finally, this overall efficiency decomposition presupposes that a strongly disposable *CRS* technology is a meaningful production model for the evaluated organization. If this is not the case, then another technology can be selected to provide the basis for an analogous, but simplified decomposition, since one or more of its components equal unity (Färe *et al.* (1994, pp. 81–82)). This remark can be linked to central concepts from the management control literature regarding responsibility centres in decentralized organizations. Depending on the autonomy to take decisions and assume operational risks, the management literature refers to (i) revenue, (ii) cost, (iii) profit and (iv) investment centres (e.g. Kaplan and Atkinson, 1998) Without exploring all these differences, it is clear that managers in cost centres are responsible for the discretional costs they decide upon and their performance assessment depends on reported cost savings, while in profit centres managers take decisions concerning both inputs and outputs and their performance depends on the profits generated. Investment centres represent an extension of profit centres whereby the accent is put on the capacity to generate profits in relation to the fixed assets deployed. It is conceivable that different responsibility centres have different needs in terms of the above decompositions, explaining the redundancy of some components.

III. Economic and Technical Capacity Utilization Concepts

Different notions of capacity exist in the literature. Specifically, it is customary to distinguish between technical (engineering) and economic (cost) capacity concepts (see, e.g. Johansen (1968) and Nelson (1989)).[17] We first treat the economic concepts using a cost frontier approach, and then the technical or engineering notion.

Traditionally, there are three basic ways of defining a cost-based notion of capacity (Nelson, 1989). The purpose of each is to isolate the short run excessive or inadequate utilization of existing fixed inputs (e.g. capital stock). The first notion of potential output is defined in terms of the output produced at short run minimum average total cost, given existing plant and factor prices (advocated by Hickman (1964), among others). It stresses the need to exploit the short run technology and the shape of the average total cost function is determined by the law of diminishing returns. The second definition corresponds to the output at which short and long run average total costs curves are tangent (following, e.g. Segerson and Squires (1990)). This is also the intersection point of short and long run expansion paths, giving this notion a particular theoretical appeal. Both notions coincide under *CRS*, since minimum of short and long run average total costs are tangent to one another. In fact, there are two variations of this tangency point notion depending on which variables one assumes to be decision variables. One notion

[16] Bogetoft *et al.* (2006) discuss how to measure allocative efficiency while maintaining technical inefficiency, which is relevant when it is easier to introduce reallocations than improvements of technical efficiency.

[17] Briec *et al.* (2010) show that it is possible to develop dual capacity measures for the case of other objective functions using nonparametric technologies: e.g. profit maximization (following Squires (1987)). The case of revenue maximization (Segerson and Squires, 1995) remains to be developed.

assumes that outputs are constant and determines optimal variable and fixed inputs. Another notion assumes that fixed inputs cannot adjust, but outputs, output prices and fixed input prices do adjust. A third definition of economic capacity, advocated by Cassels (1937) and Klein (1960), among others, considers the output determined by the minimum of the long run average total costs. It has been little used, however, probably to avoid confusion with the notion of scale economies.

For single output technologies, a capacity utilization measure can be expressed in terms of the ratio between actual output and the optimal output corresponding to the capacity notion, in which case it is termed a primal measure. Alternatively, it can be phrased in terms of the costs due to the input fixity, in which case it is labelled a dual measure. For multi-output technologies, dual measures are used most often, though Segerson and Squires (1990) have formulated some proposals to arrive at primal capacity utilization measures. This contribution focuses on dual measures in a multiple output context.[18]

To implement these cost-based notions of capacity utilization using nonparametric, deterministic frontier technologies, we summarise the possibilities.[19] One option is to select current, observed costs as a point of comparison. The resulting capacity utilization measures then compare observed costs to the reference points in the decomposition corresponding to the preferred economic capacity notion. Another option is to compare these reference points to the long run optimal costs under *CRS*, i.e. the endpoint of the traditional and the Seitz-inspired decomposition. If one takes inefficiency seriously, then starting off from the current situation seems the most natural way of defining a meaningful decomposition. But, this immediately raises the question on where to start calling inefficiency a matter of an inadequate utilization of fixed inputs. Recall that the traditional literature on capacity utilization assumes cost minimization throughout and ignores technical inefficiency altogether. Therefore, which point of comparison to use when defining measures of capacity utilization remains an open question. We return to this issue in the next section.

We first characterize the above three economic capacity notions, one of which has two variants, in a multiple output context in the following definition.

Definition 5: Reference points of economic capacity notions in the multiple output case are defined as the quantities and prices corresponding to the following:

(1) Minimum of short run total cost function $C(y, w^v, x^f|V) : C(y, w^v, x^f|C)$.
(2) Tangency cost with modified fixed inputs

$$C^{\text{tang1}}(y, w^{f*}|V):$$
$$C^{\text{tang1}}(y, w^{f*}|V) = C(y, w|V) = C(y, w^v, x^{f*}|V)$$

(3) Tangency cost with modified outputs

$$C^{\text{tang2}}(y(p, w^f, x^f), w, x^f|V):$$
$$C^{\text{tang2}}(y(p, w^f, x^f), w, x^f|V) = C(y(p, w^f, x^f), w|V) = C(y(p, w^f, x^f), w^v, x^f|V).$$

(4) Minimum of long run total cost function $C(y, w|V)$: $C(y, w|C)$,

where x^{f*} represents optimal fixed inputs, p a vector of output prices ($p \in \mathbb{R}^m_+$), and $y(p, w^f, x^f)$ represents outputs that have been adjusted in terms of given output prices, fixed input prices and the given fixed inputs.

First, the minimum of the single output short run average total cost function can be determined indirectly in the multiple output case by solving for a variable cost function relative to a *CRS* technology ($VC(y, w^v, x^f|C)$), and simply adding observed fixed costs ($FC = w^f x^f$). The resulting short run total cost function $C(y, w^v, x^f|C)$ $(= VC(y, w^v, x^f|C) + FC)$ offers the reference point for this capacity notion.

Second, the tangency point between short and long run costs can also be estimated using non-parametric cost frontiers. One can actually envision two types of tangency points depending on which variables one assumes to be decision variables. One tangency cost notion assumes that outputs remain constant and then determines optimal variable and fixed inputs ($C^{\text{tang1}}(y, w, x^{f*}|V)$). This can be solved indirectly by minimizing a long run total cost function $C(y, w|V)$ yielding optimal fixed inputs (x^{f*}). By definition, the short run total cost function with fixed inputs equal to these optimal fixed inputs ($FC(y, w^v, x^{f*}|V)$) yields exactly the same solution in terms of optimal costs and optimal variable inputs ($C(y, w^v, x^{f*}|V) = VC(y, w^v, x^{f*}|V) + FC(y, w^v, x^{f*}|V)$). Hence, the optimal solution for $C(y, w|V)$ generates the tangency point we are looking for.

[18] There is little agreement on how to define capacity utilization measures: some define it as a ratio of observed to 'optimal' costs, while others define it the reverse way (see, e.g. Segerson and Squires (1990)).
[19] Note that Coelli *et al.* (2002) define an alternative ray economic capacity measure using nonparametric frontiers that involves short-run profit maximization whereby the output mix is held constant. Though this notion has some appeal, it is rarely applied and we simply note that it does not coincide with any of the traditional capacity notions.

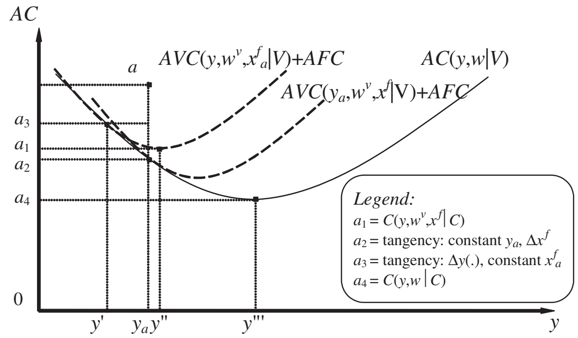

Fig. 2. Different notions of cost-based capacity utilization

Another tangency point, favoured by Nelson (1989, p. 277) and analyzed in detail in Briec *et al.* (2010), assumes that fixed inputs cannot be adjusted in the short term, but that outputs, output prices ($p \in \mathbb{R}^m_+$) and fixed input prices are adjustable such that installed capacity is utilized *ex post* at a tangency cost level ($C^{\text{tang2}}(y(p, w^f, x^f), w, x^f|V)$). Though one may object that outputs are assumed to be exogenous in a competitive cost minimization model, this tangency notion offers a useful reference point, since it retrospectively indicates the output quantities and prices as well as the fixed input prices at which existing fixed inputs would have been optimally utilised.[20] For an arbitrary observation, this tangency cost level may imply an output level ($y(p, w^f, x^f)$) below or above current outputs. Optimal costs at this tangency point are determined by solving for each observation a nonlinear system of inequalities (Briec *et al.*, 2010).

Third, the minimum of long run average total costs can be easily determined indirectly by solving for a long run total cost function defined relative to a *CRS* technology ($C(y, w|C)$). Since $OE_i(x, y, w|C)$, the ultimate point of comparison in existing static decompositions, is also defined as a ratio of $C(y, w|C)$ to observed costs (Definition 1), this amounts to simply reinterpreting the existing decompositions as measures of capacity utilization.

It is perhaps illuminating to illustrate these different economic capacity notions in the single output case in Fig. 2. For simplicity, smooth average cost functions are drawn, though also piece-wise linear approximations could be used corresponding to the nonparametric technologies employed in this contribution. The evaluated observation (a) is situated well above all curves reflecting an initial mix of technical, allocative and other inefficiencies. As the decomposition is input-oriented and holds outputs constant, the observation is vertically projected by minimizing costs according to the different notions. The figure depicts three average cost functions to illustrate all the above capacity notions: one long run cost function and two short-run cost functions. One short-run cost function traces the minimal short run average total costs for a level of fixed inputs equal to observation a ($SRATC(y, w^v, x^f_a|V) = AVC(y, w^v, x^f_a|V) + AFC$), while the other indicates the minimal short run average total costs for output levels corresponding to the same observation a $SRATC(y_a, w^v, x^f|V) = AVC(y_a, w^v, x^f|V) + AFC$. Cost level a_1 corresponds to the minimum of the short run average total cost function allowing for the optimal capital stock given current output levels ($C(y, w^v, x^f|C)$). The first tangency cost notion C^{tang1} $(y, w, x^{f*}|V)$ yields a cost level a_2 by determining optimal fixed inputs while maintaining current output levels. The second tangency cost notion C^{tang2} $(y(p, w^f, x^f), w, x^f|V)$ requires a cost level a_3 to produce an output y' (lower than y_a) with given fixed inputs. Finally, the minimum of long run average total costs ($C(y, w|C)$) is represented by cost level a_4. This would imply an output y''' (above y_a). Note that on a *CRS* technology, all economic capacity notions coincide.

Johansen (1968) pursued a technical approach focusing on a plant capacity notion.[21] Plant capacity is defined as the maximal amount that can be produced per unit of time with existing plant and equipment without restrictions on the availability of variable inputs. This capacity notion clearly takes an engineering perspective and, unlike economic capacity notions, it is not based on optimizing behaviour. Färe *et al.* (1989) and (see, e.g. Färe *et al.* (1994, Section 10.3)) include this notion into a frontier framework using output efficiency measures. Though comparability with the economic notions would be facilitated using an input orientation, such definitions are not available in the literature.[22] Therefore, the original output orientation is maintained.

[20] Though strictly speaking transgressing our framework, multiple divisions within an organization may, for instance, make such output adjustments among units in terms of respective installed capacities and their optimal utilization and eventually shut down temporarily redundant units.
[21] Johansen (1968) also proposes a synthetic capacity concept as the maximal output with existing plant and equipment while accounting for the restricted availability of variable inputs. This corresponds to technical efficiency. Since the latter notion is already part of current efficiency taxonomies, this synthetic capacity concept is ignored.
[22] Unless one would be settling for an input efficiency measure defined on the fixed input dimensions only. But we believe this contrasts too much with the focus on variable inputs in the economic capacity concepts.

An output-oriented measure of plant capacity utilization requires solving an output efficiency measure relative to both a standard technology and the same technology without restrictions on the availability of variable inputs. Plant capacity utilization in the outputs ($PCU_o(x, x^f, y)$) is defined as

$$PCU_o(x, x^f, y) = \frac{DF_o(x, y)}{DF_o(x^f, y)} \tag{10}$$

where $DF_o(x, y)$ and $DF_o(x^f, y)$ are output efficiency measures relative to technologies including respectively excluding the variable inputs. Defining both technologies, let the output set associated with technology S denote all output vectors $y \in \mathbb{R}^m_+$ that can be obtained from a given input vector $x \in \mathbb{R}^n_+$: $P(x) = \{y|(x, y) \in S\}$, and let $P(x^f) = \{y|(x^f, y) \in S\}$. Now we can define $DF_o(x, y) = \max\{\theta|\theta \geq 1, (\theta, y) \in P(x)\}$, and $DF_o(x^f, y) = \max\{\theta|\theta \geq 1, \ (\theta y) \in P(x^f)\}$. Note that $PCU_o(x, x^f, y) \leq 1$, since $1 \leq DF_o(x, y) \leq DF_o(x^f, y)$.

IV. Extending Static Efficiency Decompositions with Capacity Utilization Measures

To integrate some notion of capacity utilization into the existing static efficiency decompositions, we make the fundamental choice to start looking for improvements from the initial observation via the short run decomposition (conditioned by some input fixity) first, and then to move along the lines indicated by the long rung decomposition (where all inputs are variable). The transition term connecting both static decompositions is then linked to a notion of capacity utilization. In this way, we manage to achieve two things: (i) connect short and long run decompositions (while respecting basic duality relations), and (ii) integrate a kind of capacity utilization notion into a framework basically aimed at measuring relative performance.

Decompositions using an economic capacity concept

Our two proposals basically add another ratio of overall efficiency measures to the Seitz (1970, 1971) decomposition (DEC2) discussed above. We label this ratio of long to short run overall efficiency components a measure of dual capacity utilization. In contrast to traditional capacity utilization measures, it has a relative performance interpretation and is a key component for the integration of dual capacity utilization measures into the static efficiency decomposition. The different capacity utilization notions then differ to the extent that they eventually subsume additional components into their definition. Therefore, these extended decompositions are only partially independent of the type of economic capacity notion one prefers.

This dual capacity utilization component can be positioned before or after the cost based scale component ($CSCE_i(x, y, w)$). When positioned before $CSCE_i(x, y, w)$, the dual capacity utilization is measured relative to VRS technologies. When positioned behind $CSCE_i(x, y, w)$, the latter is evaluated using short run cost functions and the dual capacity utilization is measured relative to CRS technologies.

We first develop our two basic proposals. Next, we verify in great detail how the previous dual capacity utilization measures can be implemented within this framework. Finally, we relate some components of both new decompositions to one another and discuss the possibility to obtain additional primal information on capacity utilization.

The first extended dual decomposition (hence EDEC) is defined as follows:

$$\text{(EDEC1)} \ \ OE_i(x, y, w|C) = OE_i^{SR}(x, y, w|V) \cdot DCU_i^{SR}(x, y, w|V) \cdot CSCE_i(x, y, w)$$

where $OE_i^{SR}(x, y, w|V) = TE_i^{SR}(x, y) \cdot STE_i^{SR}(x, y) \cdot AE_i^{SR}(x, y, w|V)$. Furthermore, we have

$$\begin{aligned}
TE_i^{SR}(x, y) &= DF_i^{SR}(x, y|V, W) \\
STE_i^{SR}(x, y) &= DF_i^{SR}(x, y|V, S)/DF_i^{SR}(x, y|V, W) \\
AE_i^{SR}(x, y, w) &= OE_i^{SR}(x, y, w|V)/DF_i^{SR}(x, y|V, S) \\
DCU_i^{SR}(x, y, w|V) &= OE_i(x, y, w|V)/OE_i^{SR}(x, y, w|V) \ \text{ and} \\
CSCE_i(x, y, w) &= OE_i(x, y, w|C)/OE_i(x, y, w|V)
\end{aligned}$$

This identity includes a short run dual capacity-related term ($DCU_i^{SR}(x, y, w|V)$) and a long run scale term ($CSCE_i \ (x, y, w)$). Note that $OE_i^{SR}(x, y, w|V) = VC(y, w^v, x^f|V)/w^v x^v$ to maintain duality with $DF_i^{SR}(x, y|V, S)$. Since $OE_i(x, y, w|V) \lesseqqgtr OE_i^{SR}(x, y, w|V)$, clearly $DCU_i^{SR}(x, y, w|V) \lesseqqgtr 1$, while all other terms of the identity are bounded above by unity. Thus, $DCU_i^{SR} \ (x, y, w|V)$ measures the relative performance of long run cost minimization compared to the short run cost minimization. In fact, it is trivial to show that $DCU_i^{SR}(x, y, w|V)$ boils down to a ratio of the overall efficiency in fixed inputs solely to the overall efficiency in variable inputs

$$DCU_i^{SR}(x, y, w|V) = \frac{FC(y, w^v, x^{f*}|V)/w^f x^f}{VC(y, w^v, x^f|V)/w^v x^v}$$

A second extended dual decomposition is structured in the following identity:

$$\text{(EDEC2)} \quad OE_i(x,y,w|C) = OE_i^{SR}(x,y,w|V) \cdot CSCE_i^{SR}(x,y,w) \cdot DCU_i(x,y,w|C)$$

where $OE_i^{SR}(x,y,w|V)$ is as defined before, while

$CSCE_i^{SR}(x,y,w) = OE_i^{SR}(x,y,w|C)/OE_i^{SR}(x,y,w|V)$ and $DCU_i(x,y,w|C) = OE_i(x,y,w|C)/OE_i^{SR}(x,y,w|C)$

It includes a short run scale term ($CSCE_i^{SR}(x,y,w)$) and a long run dual capacity term ($DCU_i(x,y,w|C)$). Again all components are bounded above by unity, except $DCU_i\,(x,y,w|C) \lesseqqgtr 1$ (since $OE_i(x,y,w|C) \lesseqqgtr OE_i^{SR}(x,y,w|C)$).

This approach allows for some interesting links between the components of these two extended decompositions. For instance, the short and long run notions of scale efficiency and economic capacity utilization can be straightforwardly related to one another.[23]

To be more explicit, we discuss the potential integration of the different economic capacity notions (Definition 5) in (EDEC1) and (EDEC2) in full detail. Starting with the first economic capacity notion, it is easily fitted into (EDEC2) since the minimum of the short run cost function is part of the short-run overall efficiency (i.e. $OE_i^{SR}(x,y,w|C)$), which itself is part of the numerator of $CSCE_i^{SR}(x,y,w)$.

Second, both tangency cost notions of capacity require some elaboration. On the one hand, the notion of tangency cost at current output levels can be straightforwardly included in (EDEC1) because the numerator of $DCU_i\,(x,y,w|V)$ contains a tangency point at the long run cost function under VRS as part of its overall efficiency component in the numerator (i.e. $OE_i(x,y,w|V)$). On the other hand, the notion of tangency costs at current fixed inputs can replace the first component ($OE_i^{SR}\,(x,y,w|V)$) in both (EDEC1) and (EDEC2). To be concrete, (EDEC1) can be rewritten as

$$\begin{aligned}\text{(EDEC1}'\text{)} \quad & OE_i(x,y,w|C) \\ &= OE_i^{SR}(x, y(p,w^v,x^f), w|V) \\ &\quad \cdot DCU_i(x, y(p,w^v,x^f), y, w|V) \cdot CSCE_i(x,y,w)\end{aligned}$$

where $DCU_i(x,y(p,w^v,x^f),y,w|V) = OE_i(x,y,w|V)/OE_i^{SR}\,(x,y(p,w^v,x^f),w|V)$ and $OE_i^{SR}(x,y(p,w^v,x^f),w|V) = C(y(p,w^v,x^f),w|V)/w^v x^v$. In a similar fashion, (EDEC2) can be transformed into

$$\begin{aligned}\text{(EDEC2}'\text{)} \quad & OE_i(x,y,w|C) \\ &= OE_i^{SR}(x, y(p,w^v,x^f), w|V) \\ &\quad \cdot CSCE_i^{SR}(x, y(p,w^v,x^f), w) \\ &\quad \cdot DCU_i(x, y(p,w^v,x^f), y, w|C)\end{aligned}$$

where $CSCE_i^{SR}(x,y(p,w^v,x^f),w) = OE_i^{SR}(x,y(p,w^v,x^f),w|C)/OE_i^{SR}\,(x,y(p,w^v,x^f),w|V)$ and $DCU_i\,(x,y(p,w^v,x^f),y,w|C) = OE_i(x,y,w|C)/OE_i^{SR}(x,y(p,w^v,x^f),w|C)$. Note that $OE_i^{SR}(x,y(p,w^v,x^f),w|C) = C(y(p,w^v,x^f),w|C)/w^v x^v$.

Remark that in both (EDEC1′) and (EDEC2′) the component measures combining different output levels need not be smaller or equal to unity since the output level at tangency cost need not correspond to the output level of the evaluated observation. Note also that a way to further decompose $OE_i^{SR}(x,y(p,w^v,x^f),w|V)$ in (EDEC1′) and (EDEC2′) into its technical and allocative components (as in (EDEC1′) and (EDEC2′)) is available in Briec *et al.* (2010).

Third, as alluded to before, one can straightforwardly integrate the notion of minimal long run average total costs. Since $OE_i(x,y,w|C)$ is part of the last term in (EDEC1) and (EDEC2) (i.e. the numerator in $CSCE_i(x,y,w)$ and $DCU_i(x,y,w|C)$ respectively), this amounts to re-interpret existing decompositions as measures of capacity utilization.

To save some space, graphical illustrations of (EDEC1) and (EDEC2) are made available in Appendix A.[24]

[23] On the one hand, the link between both scale efficiency terms is simply the ratio of capacity terms

$$CSCE_i^{SR}(x,y,w) = CSCE_i(x,y,w) \cdot DCU_i(x,y,w|V)/DCU_i(x,y,w|C)$$

where the ratio of capacity notions is an adjustment factor that can be smaller, equal or larger than unity. On the other hand, the link between both economic capacity utilization notions is made by the scale terms as follows:

$$DCU_i(x,y,w|C) = DCU_i(x,y,w|V) \cdot CSCE_i(x,y,w)/CSCE_i^{SR}(x,y,w)$$

where this ratio of scale terms also forms an adjustment factor that can be smaller, equal or larger than unity.

[24] Appendices are available on the web site of the journal.

Decompositions using a technical capacity concept

When prices are unavailable or unreliable (for instance, in the public sector), it is useful to have a technical capacity concept to avoid conflating inefficiencies and differences in capacity utilization. By analogy with the extended decompositions based on an economic capacity concept, we develop two more decompositions, though these are output-oriented.

The first extended primal decomposition includes similar to (EDEC1) a short run capacity term and a long run scale term

$$\text{(EDEC3)}\quad OTE_o(x,y) = TE_o(x^f,y)\cdot STE_o(x^f,y) \cdot PCU_o(x,x^f,y|V)\cdot SCE_o(x,y)$$

where

$$TE_o(x^f,y) = DF_o(x^f,y|V,W)$$
$$STE_o(x^f,y) = DF_o(x^f,y|V,S)/DF_o(x^f,y|V,W)$$
$$PCU_o(x,x^f,y|V) = DF_o(x,y|V,S)/DF_o(x^f,y|V,S) \quad \text{and}$$
$$SCE_o(x,y) = DF_o(x,y|C,S)/DF_o(x,y|V,S)$$

Note that the traditional primal decomposition is similar to $OTE_i(x,y)$ (DEC1), but then using output-oriented rather than input-oriented efficiency measures. Since output-oriented efficiency measures are defined to be larger or equal to unity, all components of this decomposition are also larger or equal to unity, except the capacity term that is smaller or equal to unity. Note that $TE_o(x^f,y)$ and $STE_o(x^f,y)$ are defined at full plant capacity outputs, while $SCE_o(x,y)$ is defined with respect to observed outputs. In this respect, this decomposition bears some resemblance with the one based upon the tangency cost concept with given fixed inputs but adjusted outputs.

The second primal decomposition is similar to (EDEC2) and includes instead a long run capacity term and a short run scale term

$$\text{(EDEC4)}\quad OTE_o(x,y) = TE_o(x^f,y)\cdot STE_o(x^f,y) \cdot SCE_o^{SR}(x,y)\cdot PCU_o(x,x^f,y|C)$$

where $TE_o(x^f,y)$ and $STE_o(x^f,y)$ are defined as before, while

$$SCE_o^{SR}(x^f,y) = DF_o(x^f,y|C,S)/DF_o(x^f,y|V,S) \quad \text{and}$$
$$PCU_o(x,x^f,y|C) = DF_o(x,y|C,S)/DF_o(x^f,y|C,S)$$

Again all components, except the capacity component, are larger or equal to unity. Now, $TE_o(x^f,y)$, $STE_o(x^f,y)$ and $SCE_o^{SR}(x^f,y)$ are defined at full plant capacity outputs.

As in the case of the extended dual decompositions above, one can link the short and long run notions of scale efficiency and technical capacity utilization to one another.[25]

V. Empirical Illustration

To illustrate the ease of implementing the frameworks developed in this contribution, the extended decompositions of overall efficiency (EDEC1) to (EDEC4) are computed for a small sample of 16 Chilean hydro-electric power generation plants observed on a monthly basis (Atkinson and Dorfman, 2009). We limit ourselves to the observations for the year 1997 and specify an inter-temporal frontier resulting in a total of 192 units. Chile was one of the first countries deregulating its electricity market and hydro-power was a dominant source of energy during the 1990s (Pollitt, 2004)). Note that the role of hydro-power has changed during the deregulation period in that demand growth has started outpacing reserve capacity triggering questions about supply security (e.g. Bye *et al.*, 2008)).

There is one output quantity (electricity generated), the price per unit of output, and the prices and quantities of three inputs: labour, capital and water. Except for the fixed input capital, the remaining flow variables are expressed in physical units. Prices are in current Chilean pesos. Table 1 presents basic descriptive statistics for the inputs and the single output for the year 1997. Observe that the minimum price for water is zero, which corresponds to the power plants located on a river (run-of-river plants). Note that expression (3) allows for semi-positive prices. While differences in dimensionality of the cost

[25] On the one hand, the link between both scale efficiency terms is simply the ratio of capacity terms

$$SCE_o^{SR}(x^f,y) = SCE_o(x,y)\cdot PCU_o(x,x^f,y|V)/PCU_o(x,x^f,y|C)$$

where the ratio of capacity notions forms an adjustment factor that can be smaller, equal or larger than unity. On the other hand, the link between both primal capacity utilization notions is provided by the scale terms as follows:

$$PCU_o(x,x^f,y|V) = PCU_o(x,x^f,y|C)\cdot SCE_o^{SR}(x^f,y)/SCE_o(x,y)$$

where also this ratio of scale terms offers an adjustment factor that can be smaller, equal or larger than unity.

Table 1. Descriptive statistics for 1997

Variable	Trimmed mean[a]	Minimum	Maximum
Output (thousands of kWh)	46.79	0.40	353.70
Variable input (billions of m^3 of water)	126.80	0.49	1347.47
Variable input (No. of workers)	15.62	2.00	52.86
Fixed input (billions)	0.47	0.04	5.98
Output price (per kWh)	12.94	11.31	13.70
Price of water (per m^3 of water)	4.17	0.00	47.27
Price of labour (millions per worker)	1.26	1.23	1.28
Price of capital (estimated cost of capital)	0.70	0.63	0.77

Note: [a]10% trimming level.

function have a clear impact on cost levels, they need not have an impact on efficiency ratios (e.g. overall efficiency). For the reservoir plants, the price of water equals the marginal cost of fossil-fuelled generation. More details on the data are available in Atkinson and Dorfman (2009).[26]

Computing the extended decompositions of overall efficiency (EDEC1) to (EDEC4) requires solving a series of optimization models, since for each observation in the sample all components must be determined using a separate mathematical program. Most of the nonparametric frontier models used in this contribution have already appeared in the literature (Färe *et al.*, 1994; Ray, 2004). Therefore, to save some space, details on the specifications of the different efficiency measures and cost functions are made available in Appendices B and C.

To respect the multiplicative nature of the decompositions, Table 2 reports basic geometric mean results of the efficiency decompositions (EDEC1) to (EDEC4) for the complete sample (second column) as well as for run-of-river (third column) and reservoir plants (fourth column). To facilitate comparisons among decompositions, all components of (EDEC3) and (EDEC4) have been inverted. From decompositions (EDEC1) and (EDEC2), one observes that the cost efficiency level of these power plants is certainly low on average, since the frontier costs are only 22% of observed total costs.[27] In terms of its components, it is clear that a prominent problem comes from the management of the variable inputs since $OE_i^{SR}(x, y, wV)$ is lowest in both decompositions. Continuing the analysis of the common components, cost-based scale and allocative inefficiencies ($CSCE_i$ (x, y, w) and $AE_i^{SR}(x, y, w|V)$) are also important problems since they are slightly more acute than technical inefficiency ($TE_i^{SR}(x, y)$). Note that congestion ($STE_i^{SR}(x, y)$), as a special case of technical efficiency, plays a minor but nonnegligible role (about 6%).

Now we focus on the results for run-of-river and reservoir plants and test for any significant difference in efficiency distribution using the Li (1996) test.[28] Overall, reservoir plants appear as significantly more cost efficient than run-of-river plants, although their cost efficiency level does not reach the 30% level. This advantage of reservoir plants is based on significant differences in the dual capacity utilization coefficients ($DCU_i(x, y, w|V)$ and $DCU_i(x, y, w|C)$) and in their cost-based scale efficiency ($CSCE_i(x, y, w)$ and $CSCE_i^{SR}(x, y, w)$), the only two components that differ between (EDEC1) and (EDEC2). Run-of-river plants are better in terms of the structural inefficiency caused by the congestion of some inputs ($STE_i^{SR}(x, y)$), while differences among the other components are insignificant. Summing up, given their bigger size reservoir plants require more capital investments than run-of-river plants, but they have a managerial advantage in terms of cost-based scale efficiency and dual capacity utilization, since their flow of water depends less on hydrological conditions and seasonal weather variations.

From a primal perspective, (EDEC3) and (EDEC4) reveal that the production of outputs could be substantially increased to reach the frontier. Focusing on the common components in both decompositions, technical efficiency is very prominent, while congestion is almost negligible (just 3.6%).[29] Turning now to

[26] We maintain all observations rather than opting for a preliminary screening looking for any potential outliers.
[27] Atkinson and Dorfman (2009) also found considerable differences in allocative and technical inefficiencies among plants. Their efficiency levels are higher because they allow for productivity change over time as well as flexible returns to scale.
[28] The nonparametric Li (1996) test statistic compares two unknown distributions making use of kernel densities. It is rather widely used in the frontier estimation literature. Figures of the densities entering this statistic that turn out to be significantly different between run-of-river plants and reservoir plants are plotted in Appendix D.
[29] In this empirical illustration, since there is only a single output, weak and strong output disposability coincide. Therefore, we have specified weak disposability in the inputs for these output-oriented decompositions.

Table 2. Geometric mean for the efficiency decompositions (EDEC1–EDEC4)

		Complete sample (192)[a]	Run-of-river plants (132)[a]	Reservoir plants (60)[a]	Li (1996) test*
EDEC1	$OE_i\ (x, y, w \mid C)$	0.2227	0.2004	0.2810	*
	$OE_i^{SR}(x, y, w \mid V)$	0.4291	0.4224	0.4443	
	$DCU_i\ (x, y, w \mid V)$	0.8359	0.7707	0.9993	*
	$CSCE_i\ (x, y, w)$	0.6210	0.6156	0.6329	*
	$TE_i^{SR}(x, y)$	0.6903	0.6833	0.7059	
	$STE_i^{SR}(x, y)$	0.9429	0.9710	0.8839	*
	$AE_i^{SR}(x, y, w \mid V)$	0.6593	0.6366	0.7122	
EDEC2[b]	$OE_i\ (x, y, w \mid C)$	0.2227	0.2004	0.2810	*
	$OE_i^{SR}(x, y, w \mid V)$	0.4291	0.4224	0.4443	
	$DCU_i\ (x, y, w \mid C)$	0.8329	0.7889	0.9384	*
	$CSCE_i\ (x, y, w)$	0.6232	0.6014	0.6739	*
	$TE_i^{SR}(x, y)$	0.6903	0.6833	0.7059	
	$STE_i^{SR}(x, y)$	0.9429	0.9710	0.8839	*
	$AE_i^{SR}(x, y, w \mid V)$	0.6593	0.6366	0.7122	
EDEC3	$OTE_o(x, y)$	0.6317	0.6931	0.5151	*
	$TE_o(x^f, y)$	0.3014	0.3219	0.2608	
	$STE_o(x^f, y)$	0.9639	0.9722	0.9457	*
	$PCU_o(x, x^f, y \mid V)$	2.4178	2.4077	2.4400	
	$SCE_o(x, y)$	0.8994	0.9199	0.8560	*
EDEC4[b]	$OTE_o(x, y)$	0.6317	0.6931	0.5151	*
	$TE_o(x^f, y)$	0.3014	0.3219	0.2608	
	$STE_o(x^f, y)$	0.9639	0.9722	0.9457	*
	$PCU_o(x, x^f, y \mid C)$	3.4815	3.2093	4.1645	*
	$SCE_o^{SR}(x^f, y)$	0.6246	0.6901	0.5015	*

Notes: [a]The values between parentheses represent the number of units for each of the two plant types.
[b]To facilitate comparisons the order of the components of (EDEC2) and (EDEC4) follows the order of the decompositions (EDEC1) and (EDEC3). Components of (EDEC3) and (EDEC4) have been inverted to be situated below unity. Results are presented in terms of geometric means to make sure the multiplication of all components yields the original coefficient to be decomposed.
*According to the Li (1996) test, coefficients for reservoir plants are statistically different from the coefficients for run-of-river plants at the 99% confidence level.

the components that differ among decompositions, one can note that scale inefficiencies and plant capacity utilization show rather important differences in magnitude. While scale inefficiencies are small in (EDEC3), these are substantial in (EDEC4). Referring to the other component, plant capacity utilization has a more prominent role in (EDEC4).

After this general picture, we focus on the components depending on the nature of the plants. (EDEC3) and (EDEC4) show that run-of-river plants are significantly more efficient that reservoir plants, although they have a nonnegligible amount of inefficiency to be fixed (more than 30%). The differences in efficiency in favour of run-of-river plants are the scale and the structural components, while the primal capacity utilization favours the reservoir plants (in both decompositions, but only significantly so in (EDEC4)). The differences in the remaining technical efficiency component is nonsignificant.

Comparing these results with the situation of (EDEC1) and (EDEC2), the differences have the same sign for the structural and the capacity utilization components (except for (EDEC3) where the capacity component is insignificant). This indicates that the causes provoking inefficiency are in common for both the primal or dual approaches. Thus, reservoir plants are better able to manage the use of their installed capacity though they suffer from slight input congestion. Note that the primal version of the scale component signals a lower inefficiency for the run-of-river plants, which is just the reverse in the dual version when input prices are taken into account. Thus, reservoir plants operate closer than run-of-river plants to the minimal costs, but are further from the most productive scale size. This difference between dual and primal scale efficiency exemplifies relation (7) and the impact of relative short and long run allocative efficiencies.

These results reveal the relative importance of the different components influencing the long run level of efficiency of these hydro-electric power plants. From the perspective of management control, these decompositions are a tool for assessing the operating efficiency of each power plant and to discover its

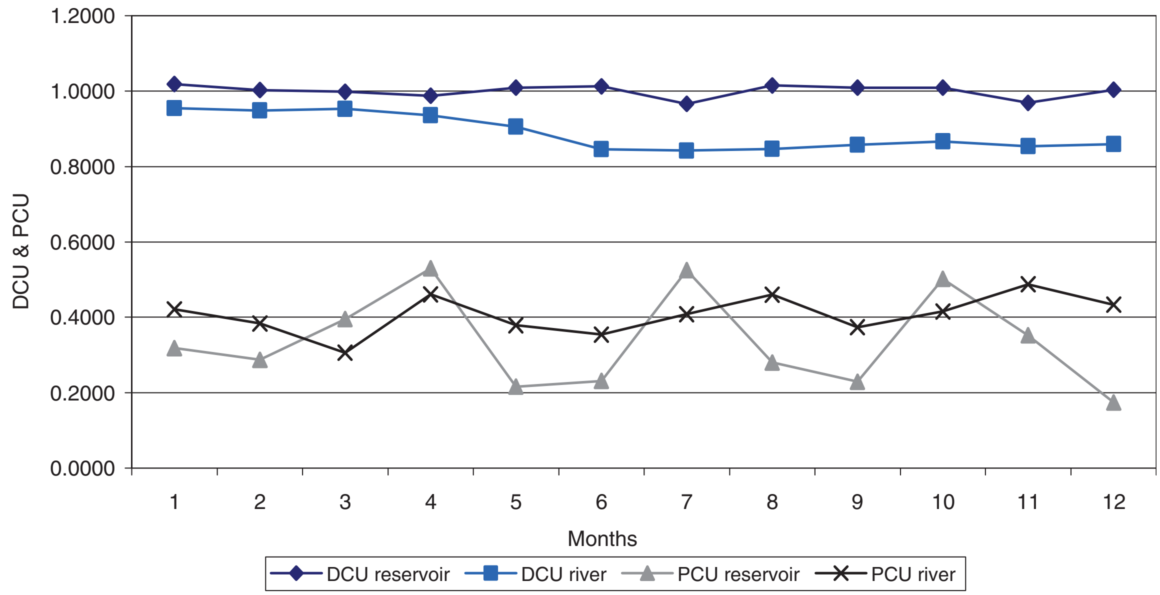

Fig. 3. Average monthly capacity components for run-of-river versus reservoir plants (DCU (EDEC1) and PCU (EDEC4))

specific strong and weak points. Managers can take advantage of these components to design actions targeting at operating with efficient cost levels.

After this general picture, we focus on the capacity components developed above in terms of the nature of the plants. In Fig. 3, we trace their variation by comparing the average monthly evolution of run-of-river versus reservoir plants in 1997. This illustrates the potential dual role of these power plants: run-of-river plants are used for base load, while reservoir plants play a role in both base load and peak periods. Comparing one cost-based notion of capacity ($DCU_i(x, y, w|V)$), given their dual role in the electricity system it is evident that reservoir plants are able to manage total and variable costs with a stable level of efficiency through the year. For run-of-river plants, one observes some seasonal variation. For a plant capacity component ($PCU_o(x, x^f, y|V)$) one typically observes a lot more seasonal variation. For run-of-river plants this simply reflects hydrological conditions: in summer (winter) times we see a substantial drop (increase) in their capacity. The strong variability of the reservoir plants illustrates the importance of their intertemporal allocation decisions in response to changes in peak demand. However, these scheduling decisions are not reflected in the cost component.

VI. Conclusions

This article has first reviewed the traditional way of defining different sources of efficiency. Having developed the ways in which both technical and economical capacity utilization concepts can be made operational, the traditional decomposition of efficiency has been extended in several ways by integrating either an economical or a technical notion of capacity utilization. An empirical illustration using a monthly panel of Chilean hydro-electric power plants demonstrates the potential of these new decomposition proposals.

This work establishes a firmer link between efficiency measurement and the traditional economic analyses of short and long run production behaviour. Of course, also the definition of identities should ideally be put to an empirical test to assess their pertinence. In our view, apart from academic empirical applications, this would imply checking the opinion of policy makers (e.g. regulators) and managers employing these frontier benchmarking tools. For instance, in incentive regulatory mechanisms (e.g. price cap regulation) the distinction between technical inefficiency and capacity utilization issues has been given insufficiently attention. This could be a topic worthy of further exploration.

One possible extension is to derive capacity notions for indirect technologies where output maximizing production is, e.g. subject to a budget constraint (Ray *et al.* (2006)) for non-parametric capacity notions in this context), or for regulated industries (e.g. Segerson and Squires (1993)). Another issue is the development of statistical test procedures for these boundary estimators to check whether, e.g. some components are significantly different from zero for a given sample, time period, and sector (along the lines of Simar and Wilson (2000)). A final extension includes the integration of these capacity terms into the productivity measurement literature. Indeed, when panel data are available, it would be useful to integrate these extended decompositions into a dynamic analysis of productivity change. A start has been made by, for instance, De Borger and Kerstens (2000) who have included the plant capacity notion into the definition of a primal Malmquist productivity index (see Zofio (2007) for a survey). Though some first steps have been taken, discrete time dual productivity indexes could probably equally benefit from the integration of economic capacity terms.[30]

Acknowledgements

The authors are grateful to the referees, W. Briec and G. Hites for most helpful comments.

References

Athanassopoulos, A. and Giokas, D. (2000) The use of data envelopment analysis in banking institutions: evidence from the Commercial Bank of Greece, *Interfaces*, **30**, 81–95.

Atkinson, S. E. and Dorfman, J. H. (2009) Feasible estimation of firm-specific allocative inefficiency through Bayesian numerical methods, *Journal of Applied Econometrics*, **24**, 675–97.

Balcombe, K., Davidova, S. and Latruffe, L. (2008) The use of bootstrapped Malmquist indices to reassess productivity change findings: an application to a sample of Polish farms, *Applied Economics*, **40**, 2055–61.

Balk, B. M. (2001) Scale efficiency and productivity change, *Journal of Productivity Analysis*, **15**, 159–83.

Banker, R., Charnes, A. and Cooper, W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, **30**, 1078–92.

Bogetoft, P., Färe, R. and Obel, B. (2006) Allocative efficiency of technically inefficient production units, *European Journal of Operational Research*, **168**, 450–62.

Bonin, J. P., Derek, C. J. and Putterman, L. (1993) theoretical and empirical studies of producer cooperatives: will ever the Twain meet, *Journal of Economic Literature*, **31**, 1290–320.

Briec, W., Kerstens, K., Prior, D. and Van de Woestyne, I. (2010) Tangency capacity notions based upon the profit and cost functions: a non-parametric approach and a comparison, *Economic Modelling*, **27**, 1156–66.

Briec, W., Kerstens, K. and Vanden Eeckaut, P. (2004) Non-convex technologies and cost functions: definitions, duality and nonparametric tests of convexity, *Journal of Economics*, **81**, 155–92.

Bye, T., Bruvoll, A. and Roar Aune, F. (2008) Inflow shortages in deregulated power markets – reasons for concern?, *Energy Economics*, **30**, 1693–711.

Cassels, J. M. (1937) Excess capacity and monopolistic competition, *Quarterly Journal of Economics*, **51**, 426–43.

Caves, R. E. (2007) In praise of the Old I.O., *International Journal of Industrial Organization*, **25**, 1–12.

Chambers, R., Chung, Y. and Färe, R. (1998) Profit, directional distance functions, and Nerlovian efficiency, *Journal of Optimization Theory and Applications*, **98**, 351–64.

Coelli, T., Grifell-Tatjé, E. and Perelman, S. (2002) Capacity utilization and profitability: a decomposition of short-run profit efficiency, *International Journal of Production Economics*, **79**, 261–78.

De Borger, B. and Kerstens, K. (2000) The Malmquist productivity index and plant capacity utilization, *Scandinavian Journal of Economics*, **102**, 303–10.

Färe, R. (1984) The existence of plant capacity, *International Economic Review*, **25**, 209–13.

Färe, R. and Grosskopf, S. (2000) Decomposing technical efficiency with care, *Management Science*, **46**, 167–8.

Färe, R., Grosskopf, S. and Kokkelenberg, E. (1989) Measuring plant capacity, utilization and technical change: a nonparametric approach, *International Economic Review*, **30**, 655–66.

Färe, R., Grosskopf, S. and Lovell, C. A. K. (1983) The structure of technical efficiency, *Scandinavian Journal of Economics*, **85**, 181–90.

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

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

Farrell, M. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*, **120A**, 253–81.

Førsund, F. and Hjalmarsson, L. (1974) On the measurement of productive efficiency, *Swedish Journal of Economics*, **76**, 141–54.

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

Fukuyama, H. and Weber, W. (1999) The efficiency and productivity of Japanese securities firms, 1988–93, *Japan and the World Economy*, **11**, 115–33.

Fuss, M., McFadden, D. and Mundlak, Y. (1978) A survey of funtional forms in the economic analysis of production, in *Production Economics: A Dual Approach to Theory and Applications*, Vol. 1 (Eds) M. Fuss and D. McFadden, North-Holland, Amsterdam, pp. 219–68.

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

Ghemawat, P. and Nalebuff, B. (1985) Exit, *Rand Journal of Economics*, **16**, 184–94.

Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S. and Stringer, K. S. (2009)

[30] Though this may well not be that easy, given that even the precise integration of scale efficiency into the Malmquist productivity index has been the source of considerable controversy (see Balk (2001) for an overview).

Best-practice benchmarking in UK higher education: new nonparametric approaches using financial ratios and profit efficiency methodologies, *Applied Economics*, **41**, 249–67.

Grifell-Tatjé, E. and Lovell, C. A. K. (2000) Cost and productivity, *Managerial and Decision Economics*, **21**, 19–30.

Grosskopf, S. (1986) The role of the reference technology in measuring productive efficiency, *Economic Journal*, **96**, 499–513.

Hickman, B. G. (1964) On a new method for capacity estimation, *Journal of the American Statical Association*, **59**, 529–49.

Hulten, C. (1986) Productivity change, capacity utilization, and the sources of productivity growth, *Journal of Econometrics*, **33**, 31–50.

Jamasb, T. and Pollitt, M. (2001) Benchmarking and regulation: international electricity experience, *Utilities Policy*, **9**, 107–30.

Johansen, L. (1968) *Production Functions and the Concept of Capacity, Namur, Recherches Récentes sur la Fonction de Production* (Collection 'Economie Mathématique et Econometrie', n° 2) [reprinted in F.R. Førsund (ed.) (1987) Collected Works of Leif Johansen, Vol. 1, North Holland, Amsterdam, pp. 359–82].

Kallio, M. and Kallio, M. (2002) Nonparametric methods for evaluating economic efficiency and imperfect competition, *Journal of Productivity Analysis*, **18**, 171–89.

Kaplan, R. S. and Atkinson, A. A. (1998) *Advanced Management Accounting*, 3rd edn, Prentice Hall, London.

Klein, L. R. (1960) Some theoretical issues in the measurement of capacity, *Econometrica*, **28**, 272–86.

Li, Q. (1996) Nonparametric testing of closeness between two unknown distribution functions, *Econometric Reviews*, **15**, 261–74.

McDonald, J. (1996) A problem with the decomposition of technical inefficiency into scale and congestion components, *Management Science*, **42**, 473–4.

Nelson, R. (1989) On the measurement of capacity utilization, *Journal of Industrial Economics*, **37**, 273–86.

Pollitt, M. G. (2004) Electricity reform in Chile: lessons for developing countries, *Journal of Network Industries*, **5**, 221–62.

Ray, S. C. (2004) *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, Cambridge University Press, Cambridge.

Ray, S. C., Mukherjee, K. and Wu, Y. (2006) Direct and indirect measures of capacity utilization: a nonparametric analysis of US manufacturing, *Manchester School*, **74**, 526–48.

Rowland, W. W., Langemeier, M. R., Schurle, B. W. and Featherstone, A. M. (1998) A nonparametric efficiency analysis for a sample of Kansas swine operations, *Journal of Agricultural and Applied Economics*, **30**, 189–99.

Segerson, K. and Squires, D. (1990) On the measurement of economic capacity utilization for multi-product industries, *Journal of Econometrics*, **44**, 347–61.

Segerson, K. and Squires, D. (1993) Capacity utilization under regulatory constraints, *Review of Economics and Statistics*, **75**, 76–85.

Segerson, K. and Squires, D. (1995) Measurement of capacity utilization for revenue-maximizing firms, *Bulletin of Economic Research*, **47**, 77–85.

Seitz, W. D. (1970) The measurement of efficiency relative to a frontier production function, *American Journal of Agricultural Economics*, **52**, 505–11.

Seitz, W. D. (1971) Productive efficiency in the steam-electric generating industry, *Journal of Political Economy*, **79**, 878–86.

Simar, L. and Wilson, P. W. (2000) Statistical inference in nonparametric frontier models: the state of the art, *Journal of Productivity Analysis*, **13**, 49–78.

Squires, D. (1987) Long-run profit functions for multi-product firms, *American Journal of Agricultural Economics*, **69**, 558–69.

Sueyoshi, T. (1999) DEA duality on returns to scale (RTS) in production and cost analyses: an occurrence of multiple solutions and differences between production-based and cost-based RTS estimates, *Management Science*, **45**, 1593–608.

Tone, K. and Sahoo, B. K. (2003) Scale, indivisibilities and production function in data envelopment analysis, *International Journal of Production Economics*, **84**, 165–92.

Zofio, J. L. (2007) Malmquist productivity index decompositions: a unifying framework, *Applied Economics*, **39**, 2371–87.