HUB RESEARCH PAPER
Economics & Management

Static Efficiency Decompositions and Capacity Utilisation: Integrating Economic and Technical Capacity Notions

Bruno De Borger, Kristiaan Kerstens, Diego Prior, Ignace Van de Woestyne

HUB RESEARCH PAPER 2009/30
SEPTEMBER 2009

STATIC EFFICIENCY DECOMPOSITIONS AND CAPACITY UTILISATION: INTEGRATING ECONOMIC AND TECHNICAL CAPACITY NOTIONS

Bruno De Borger*

Kristiaan Kerstens**

Diego Prior***

Ignace Van de Woestyne****

Abstract

Starting from existing static decompositions of overall economic efficiency on non-parametric production and cost frontiers, this paper proposes more comprehensive decompositions including several cost-based notions of capacity utilisation. Furthermore, in case prices are lacking, we develop additional decompositions of overall technical efficiency integrating a technical concept of capacity utilisation. 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 utilisation. An empirical analysis using a monthly panel of Chilean hydro-electric power plants illustrates the potential of these decomposition proposals.

Keywords: efficiency; capacity utilisation.

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

* University of Antwerp, Department of Economics, Prinsstraat 13, B-2000 Antwerp, Belgium.

** CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France. Tel: ++ 33(0)320545892, Fax: ++ 33(0)320574855, k.kerstens@ieseg.fr. Corresponding author.

*** Universitat Autònoma de Barcelona, Dpt. d'Economia de l'Empresa, E-08193 Bellaterra, Spain.

**** Hogeschool-Universiteit Brussel, Stormstraat 2, B-1000 Brussel, Belgium.

March 2007
This revision: September 2009

1. INTRODUCTION

The analysis of efficiency and productivity based on frontier specifications of technology has become a standard empirical tool serving both regulatory and managerial purposes (see, e.g, Jamasb and Pollitt (2001) for a survey of countries implementing performance based regulation in electricity). However, it is a bit surprising that many if not most applications have –often implicitly- taken a long run perspective in that 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 recognised for long, the literature has almost completely ignored the notion of capacity utilisation. 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 utilisation. 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 utilisation becomes an organisational factor. For example, Bonin, Derek and Putterman (1993) report that cooperative firms are able to maintain more stable production plans than non-cooperative firms, which is a factor that seems related with the advantages of having stable contracts with regular partners.

This paper concentrates on the development of efficiency decompositions integrating capacity utilisation using non-parametric frontier technologies. In this non-parametric approach, piecewise linear frontiers envelop the observations as tightly as possible subject to certain minimal production axioms.[1] More specifically, this paper makes -to the best of our knowledge- two contributions. First, this is the first proposal in the literature integrating different notions of capacity utilisation into a taxonomy of static efficiency concepts for non-parametric technologies.[2] Second, we integrate both primal and dual concepts of capacity into this literature on multiple output non-parametric frontiers.[3] In brief, the purpose of our contribution is to carefully disentangle between capacity utilisation and various efficiency concepts in a non-parametric

[1] Parametric frontiers utilise parametric, locally if not even globally flexible specifications with a finite number of parameters to estimate the underlying technology.

[2] In the parametric literature various productivity decompositions have been suggested to include measures of capacity utilisation (see Hulten (1986) and Morrison (1985, 1993), among others). Some productivity decompositions have been recently proposed in the non-parametric frontier literature (see below).

[3] Mainly dual multiple output concepts are known in the parametric literature (e.g., (Morrison (1985) and Squires (1987)), while primal capacity notions are difficult to estimate. Färe (1984) shows that a primal capacity notion cannot

frontier framework that allows for a coherent treatment of both primal and dual capacity notions. Already Fuss, McFadden and Mundlak (1978: 223) stressed that fundamental production axioms are of a qualitative and non-parametric nature and therefore should ideally be tested using non-parametric technologies.[4]

The paper is structured as follows. Section 2 summarises the traditional static decomposition of overall economic efficiency and some less known useful extensions. The next section introduces both economic (cost-based) and technical concepts of capacity utilisation. Section 4 extends the traditional efficiency decomposition by integrating this variety of capacity utilisation measures. These new decompositions are related to one another, with a focus on the relations between short and long run scale efficiency and capacity utilisation. In addition, decompositions of overall technical efficiency integrating a technical concept of capacity utilisation 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 6.

2. DEFINITION OF THE STATIC EFFICIENCY DECOMPOSITION

2.1. 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: $S = \{(x,y) \mid x \text{ can produce } y\}$. The input set associated with S denotes all input vectors $x \in \mathbb{R}^n_+$ capable of producing a given output vector $y \in \mathbb{R}^m_+$: $L(y) = \{x \mid (x,y) \in S\}$. 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 the following standard assumptions (see Färe, Grosskopf and Lovell (1994)):

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 $\cap_{n\in\mathbb{N}}L(y_n) \neq 0$.

L3: $L(y)$ is closed $\forall\, y \in \mathbb{R}^m_+$.

L4: $L(y)$ is a convex set $\forall\, y \in \mathbb{R}^m_+$.

---

be obtained for certain popular parametric specifications of technology (e.g., the Cobb-Douglas). Recently, however, Kirkley, Morrison Paul and Squires (2004) presented parametric primal capacity estimates.

[4] This does not preclude an eventual extension of our proposals into a parametric framework.

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.

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:

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

We next define the radial input efficiency measure as:

$$DF_i(x, y) = \min \{\lambda \mid \lambda \geq 0, (\lambda x) \in L(y)\}. \quad (2)$$

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, Grosskopf and Lovell (1994) for details).[5]

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

$$C(y,w) = \min \{wx \mid x \in L(y)\}. \quad (3)$$

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

This dual relation establishes the foundations for efficiency measurement.[6] 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:

[5] For reasons of convenience, we stick to the traditional radial input efficiency measure, i.e., the inverse of the input distance function. Recently, more general directional distance functions have been introduced to measure profit efficiency (see Chambers, Chung and Färe (1998)). Apart from the fact that these new measures lead to additive rather than multiplicative decompositions, they can be exactly related to the traditional radial efficiency measures employed in this contribution.

[6] The duality relation between input distance function and cost function is:
$C(y, w) = \min_x \{wx : D_i(x, y) \geq 1\} \quad w > 0$ and $D_i(x, y) = \min_w \{wx : C(y, w) \geq 1\} \quad x \in L(y)$.

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

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

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

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

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

The first ratio of mimimal to observed costs $C(y,w)/w \cdot x$ defines a cost efficiency component (labeled an 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 $AE(x,y,w)$ indicates the allocative efficiency, defined in a residual way.

Furthermore, since the main focus of this contribution is on establishing a link between existing efficiency decompositions and traditional capacity concepts, 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:

$$DF_i^{SR}(x,y) = \min \left\{ \lambda \mid \lambda \ge 0, (\lambda\, x^v, x^f) \in L(y) \right\}. \qquad (7)$$

Duplicating the above definitions, one can straightforwardly develop an analogous sub-vector efficiency decomposition.

2.2. 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, Grosskopf and Lovell (1983, 1985: 3-5) offer an extended efficiency taxonomy.[7] Since technologies vary in terms of underlying production assumptions (see Färe, Grosskopf and Lovell (1994)), it is useful to condition the above notation of the efficiency measure on two main assumptions: (i) the difference between

While the cost function can be obtained from the input distance function by optimizing with respect to input quantities, the input distance function can be resolved from the cost function by minimizing with respect to input prices.

[7] Other classifications include Banker, Charnes and Cooper (1984) and Førsund and Hjalmarsson (1974, 1979).

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 their proposals have become the 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.[8] Notice that this and all other extended static efficiency decompositions discussed below fundamentally start from the basic multiplicative Farrell (1957) decomposition (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:* $TE_i(x,y) = DF_i(x,y \mid V,W)$.
2) *Structural Efficiency is the quantity:* $STE_i(x,y) = DF_i(x,y \mid V,S)/DF_i(x,y \mid V,W)$.
3) *Scale Efficiency is the quantity:* $SCE_i(x,y) = DF_i(x,y \mid C,S)/DF_i(x,y \mid V,S)$.
4) *Overall Technical Efficiency is the quantity:* $OTE_i(x,y) = DF_i(x,y \mid C,S)$.
5) *Overall Efficiency is the quantity:* $OE_i(x,y,w) = C(y,w \mid C)/wx$.
6) *Allocative Efficiency is the quantity:* $AE_i(x,y,w) = OE_i(x,y,w)/OTE_i(x,y)$.

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. In fact, $TE_i(x,y)$ is traditionally evaluated relative to a *VRS* technology with *WD* using $DF_i(x,y \mid V,W)$. Second, structural efficiency ($STE_i(x,y)$) implies that production should occur 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*.[9] Finally, overall technical efficiency

[8] To simplify matters, we ignore the analysis of efficiency in non-competitive settings, leading to price inefficiencies in addition to inefficiencies in quantities (see, e.g., Färe, Grosskopf and Lovell (1994), Grifell-Tatjé and Lovell (2000), or Kallio and Kallio (2002)).

[9] In addition, it is possible to obtain qualitative information regarding 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 non-increasing returns to scale ($DF_i(x,y \mid N,S)$) and by exploiting the nestedness of technologies, one discriminates between *IRS* and *DRS* (Färe, Grosskopf and Lovell

($OTE_i(x,y)$) 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 ($OE_i(x,y,w)$) requires computing a cost function relative to a *CRS* technology with *SD* and taking a ratio of this minimal costs to actual costs. $OE_i(x,y,w)$ can be seen as the multiplicative result of $OTE_i(x,y)$ and allocative efficiency ($AE_i(x,y,w)$), defined as a residual term making up the gap between $OE_i(x,y,w)$ and $OTE_i(x,y)$. $AE_i(x,y,w)$ 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.

Notice that $OE_i(x,y,w)$ and $AE_i(x,y,w)$ imply price-dependent characterisations 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 of these components are smaller or equal to unity. These static efficiency concepts are mutually exclusive and their radial measurement yields a multiplicative decomposition:

$$OE_i(x,y,w) = AE_i(x,y,w).OTE_i(x,y) \qquad \text{(DEC1)}$$

where $OTE_i(x,y) = TE_i(x,y).STE_i(x,y).SCE_i(x,y)$ (see Färe, Grosskopf and Lovell (1985)).

This traditional static efficiency decomposition is illustrated on Figure 1 representing three technologies: one imposing *SD* and *CRS* ($L(y \mid C,S)$); one with *SD* and *VRS* ($L(y \mid V,S)$); and one with *WD* and *VRS* ($L(y \mid V,W)$). For observation g, $OE_i(x,y,w)$ 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 $AE_i(x,y,w) = 0g_6/0g_4$.

<FIGURE 1 ABOUT HERE>

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 4, since scale efficiency is based on a dual characterisation 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:*

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

(1983)): (i) *IRS* holds when $DF_i(x,y \mid C,S) = DF_i(x,y \mid N,S) \leq DF_i(x,y \mid V,S) \leq 1$; (ii) *DRS* holds when $DF_i(x,y \mid C,S) \leq DF_i(x,y \mid N,S) = DF_i(x,y \mid V,S) \leq 1$.

$CSCE_i(x,y,w)$ is a price-dependent scale efficiency term based on cost function estimates. Since $OE_i(x,y,w \mid C) \le OE_i(x,y,w \mid V)$, $CSCE_i(x,y,w) \le 1$.[10]

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

$$OE_i(x,y,w \mid C) = CSCE_i(x,y,w).OE_i(x,y,w \mid V), \qquad \text{(DEC2)}$$

whereby $OE_i(x,y,w \mid V) = TE_i(x,y).STE_i(x,y).AE_i(x,y,w \mid V)$. $CSCE_i(x,y,w)$ is labelled "economic scale efficiency" by Seitz (1970: 508), while $OE_i(x,y,w \mid V)$ is termed "economic efficiency given the scale" of operations.[11] 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.[12] This alternative decomposition is also illustrated on Figure 1. For observation g, $OE_i(x,y,w \mid C)$ 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 $AE_i(x,y,w \mid V) = 0g_5/0g_3$ 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 $CSCE_i(x,y,w)$ into its technical and allocative components:

$$CSCE_i(x,y,w) = \left[\frac{DF_i(x,y|C,S)}{DF_i(x,y|V,S)}\right] \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)}\right] = SCE_i(x,y) \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)}\right], \qquad (8)$$

Färe and Grosskopf (1985) and Färe, Grosskopf and Lovell (1994) show: $CSCE_i(x,y,w) = SCE_i(x,y) \Leftrightarrow AE_i(x,y,w \mid C) = AE_i(x,y,w \mid V)$.[13] Since $OE_i(x,y,w \mid C) \le OE_i(x,y,w \mid V) \le 1$,

[10] Identification of local returns to scale proceeds as follows. When $CSCE_i(x,y,w) = 1$, then the unit minimises costs and enjoys *CRS*. When $CSCE_i(x,y,w) < 1$, then computing a cost function relative to a non-increasing returns to scale technology ($OE_i(x,y,w \mid N)$) and knowing that $OE_i(x,y,w \mid C) \le OE_i(x,y,w \mid N) \le OE_i(x,y,w \mid V) \le 1$ (Grosskopf (1986)), the same reasoning as above applies to infer local scale economies. This procedure applies to any dual formulation.

[11] 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) or Rowland et al. (1998).

[12] To clarify matters, one could introduce the notation $AE_i(x,y,w \mid C)$ in (DEC1) to distinguish this component from the one in (DEC2).

[13] See also Sueyoshi (1999). Actually, scale efficiency in Färe and Grosskopf (1985) 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 organisations face identical input prices; and (ii) $AE_i(x,y,w \mid C) = AE_i(x,y,w \mid V)$. When input price information is available and cost function estimates are employed, however, the first of these conditions is redundant.

$[AE_i(x,y,w \mid C)/AE_i(x,y,w \mid V)] \lesseqqgtr 1$. 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:

$$OE_i(x,y,w|C) = SCE_i(x,y) \cdot \left[\frac{AE_i(x,y,w|C)}{AE_i(x,y,w|V)} \right] .TE_i(x,y).STE_i(x,y).AE_i(x,y,w|V), \qquad (9)$$

where $OE_i(x,y,w \mid V)$ contains the last three terms and eliminating the common $AE_i(x,y,w \mid V)$ term yields (DEC1).

Though Färe, Grosskopf and Lovell (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.[14] 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.

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 minimisation as a realistic goal, one would expect organisations to minimise 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.[15] But, the whole point of retrospectively benchmarking organisational performance is that organisations 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 utilisation 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.

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

[15] Bogetoft, Färe and Obel (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.

Finally, this overall efficiency decomposition presupposes that a strongly disposable *CRS* technology is a meaningful production model for the evaluated organisation. 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, Grosskopf and Lovell (1994: 81-82)). This remark can be linked to central concepts from the management control literature regarding responsibility centres in decentralised organisations. 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 (1989)) 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. For instance, for a cost centre itself scale efficiency may be little relevant, while it is of utmost importance to an investment centre.

3. ECONOMIC AND TECHNICAL CAPACITY UTILISATION 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), Morrison (1993), Nelson (1989)).[16] 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 (see Morrison (1985), Nelson (1989)). The purpose of each is to isolate the short run excessive or inadequate utilisation 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 Berndt and Morrison (1981) and 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

[16] Briec, Kerstens and Prior (2006) show that it is possible to develop dual capacity measures for the case of other objective functions using non-parametric technologies: e.g., profit maximisation (following Squires (1987)). The case of revenue maximisation (see Segerson and Squires (1995)) remains to be developed.

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 (see Morrison (1985)). One notion 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 Cassel (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 utilisation 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 (Morrison (1985)), though Segerson and Squires (1990) have formulated some proposals to arrive at primal capacity utilisation measures. This contribution focuses on dual measures in a multiple output context.[17]

To implement these cost-based notions of capacity utilisation using non-parametric, deterministic frontier technologies, we summarise the possibilities.[18] One option is to select current, observed costs as a point of comparison. The resulting capacity utilisation 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 utilisation of fixed inputs. Recall that the traditional literature on capacity utilisation assumes cost minimisation throughout and ignores technical inefficiency altogether. Therefore, which point of comparison to use when defining measures of capacity utilisation remains an open question. We return to this issue in the next section.

---

[17] There is little agreement on how to define capacity utilisation measures: some define it as a ratio of observed to "optimal" costs (e.g., Morrison (1985)), while others define it the reverse way (e.g., Segerson and Squires (1990)).

[18] Notice that Coelli, Grifell-Tatjé and Perelman (2002) define an alternative ray economic capacity measure using non-parametric frontiers that involves short-run profit maximisation whereby the output mix is held constant. Though this notion has some appeal, it is rarely applied (e.g, Pascoe and Tingley (2006)) and we simply note that it does not coincide with any of the traditional capacity notions.

We first characterise the above three economic capacity notions, one of which has 2 variants, in a multiple output context in the following definition.

Definition 3: *Reference points of economic capacity notions in the multiple output case are defined as the quantities and prices corresponding to:*

1) *Minimum of short run total cost function:* $C(y,w^{v},x^{f}\,|\,C)$.
2) *Tangency cost with modified fixed inputs:* $C^{tang1}(y,w,x^{f}{*}\,|\,V) = C(y,w\,|\,V) = C(y,w^{v},x^{f}{*}\,|\,V)$.
3) *Tangency cost with modified outputs, output prices and fixed input prices:* $C^{tang2}(y(p,w^{f},x^{f}),w,x^{f}\,|\,V) = C(y(p,w^{f},x^{f}),w\,|\,V) = C(y(p,w^{f},x^{f}),w^{v},x^{f}\,|\,V)$.
4) *Minimum of long run total cost function:* $C(y,w\,|\,C)$.

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 ($VC(y,w^{v},x^{f}\,|\,C) = \min\{w^{v}x^{v}\;|\;(x^{v},x^{f}) \in L(y\,|\,C,S)\}$), solved by a simple linear program, 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 ($C^{tang1}(y,w,x^{f}{*}\,|\,V)$). This can be solved indirectly by minimising 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 yields exactly the same solution in terms of optimal costs and optimal variable inputs ($C(y,w^{v},x^{f}{*}\,|\,V) = VC(y,w^{v},x^{f}{*}\,|\,V) + w^{f}.x^{f}{*}$). Hence, the optimal solution for $C(y,w\,|\,V)$ generates the tangency point we are looking for.

Another tangency point, favoured by Nelson (1989: 277) and analysed in detail in Briec, Kerstens and Prior (2009), 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 utilised ex post at a tangency cost level ($C^{tang2}(y(p,w^{f},x^{f}),w,x^{f}\,|\,V)$). Though one may object that outputs are assumed to be exogenous in a competitive cost minimisation 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.[19] 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 non-linear system of inequalities (see Briec, Kerstens and Prior (2009)).

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: $C(y,w \mid C) = \min \{wx \mid x \in L(y \mid C,S)\}$. Since $OE_i(x,y,w \mid C)$, the ultimate point of comparison in existing static decompositions, is also defined as a ratio of $C(y,w \mid C)$ to observed costs (see Def. 1), this amounts simply to reinterpreting the existing decompositions as measures of capacity utilisation.

It is perhaps illuminating to illustrate these different economic capacity notions in the single output case in Figure 2. For simplicity, smooth average cost functions are drawn, though also piece-wise linear approximations could be used corresponding to the non-parametric 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 minimising costs according to the different notions. The figure depicts three average cost functions to illustrate all 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 ($SRATC(y,w^v,x^f_a \mid V) = AVC(y,w^v,x^f_a \mid V) + AFC$), while the other indicates the minimal short run average total costs for output levels corresponding to the same observation a $SRATC(y_a,w^v,x^f \mid V) = AVC(y_a,w^v,x^f \mid V) + AFC$. 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 ($C(y,w^v,x^f \mid C)$). The first tangency cost notion $C^{tangency1}(y,w,x^{f*} \mid V)$ yields a cost level a_2 by determining optimal fixed inputs while maintaining current output levels. The second tangency cost notion $C^{tangency2}(y(p,w^f,x^f),w,x^f \mid V)$ 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 ($C(y,w \mid C)$) is represented by cost level a_4. This would imply an output y''' (above y_a). Notice that on a *CRS* technology, all economic capacity notions coincide.

<FIGURE 2 ABOUT HERE>

[19] Though strictly speaking transgressing our framework, multiple divisions within an organisation may, for instance, make such output adjustments among units in terms of respective installed capacities and their optimal utilisation and eventually shut down temporarily redundant units.

Johansen (1968) pursued a technical approach focusing on a plant capacity notion.[20] 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 optimising behaviour. Färe, Grosskopf and Kokkelenberg (1989) and Färe, Grosskopf and Valdmanis (1989) 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.[21] Therefore, the original output orientation is maintained.

An output-oriented measure of plant capacity utilisation 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 utilisation in the outputs ($PCU_o(x,x^f,y)$) is defined as:

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

where $DF_o(x,y)$ and $DF_o(x^f,y)$ are output efficiency measures relative to technologies including respectively excluding the variable inputs. Indeed, $DF_o\,(x, y) = \max\left\{\theta \;\middle|\; \theta \geq 1, (\theta\, y) \in P(x)\right\}$, where the output set associated with S denotes all output vectors $y \in \mathbb{R}^m_+$ that can be obtained from a given input vector $x \in \mathbb{R}^n_+$: $P(x) = \{y \mid (x,y) \in S\}$. Furthermore, $DF_o\,(x, y) = \max\left\{\theta \;\middle|\; \theta \geq 1, (\theta\, y) \in P(x^f)\right\}$, where $P(x^f) = \{y \mid (x^f,y) \in S\}$. Notice that $PCU_o(x,x^f,y) \leq 1$, since $1 \leq DF_o(x,y) \leq DF_o(x^f,y)$.

4. EXTENDING STATIC EFFICIENCY DECOMPOSITIONS WITH CAPACITY UTILISATION MEASURES

4.1. 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 utilisation. It has a relative performance interpretation and it is a key component for the integration of dual capacity

[20] 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.

[21] 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.

utilisation measures into the static efficiency decomposition. The different capacity utilisation 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 utilisation 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 utilisation 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 utilisation is measured relative to CRS technologies.

We first propose our two basic proposals. Next, we verify in great detail how the previous dual capacity utilisation 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 utilisation.

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

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

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

$$TE_i^{SR}(x,y) = DF_i^{SR}(x,y|V,W);$$

$$STE_i^{SR}(x,y) = DF_i^{SR}(x,y|V,S)\,/\,DF_i^{SR}(x,y|V,W);$$

$$AE_i^{SR}(x,y,w) = OE_i^{SR}(x,y,w|V)\,/\,DF_i^{SR}(x,y|V,S);$$

$$DCU_i^{SR}(x,y,w|V) = OE_i\,(x,y,w|V)\,/\,OE_i^{SR}(x,y,w|V); \text{ and}$$

$$CSCE_i\,(x,y,w) = OE_i\,(x,y,w|C)\,/\,OE_i\,(x,y,w|V).$$

This identity includes a short run dual capacity-related term ($DCU_i^{SR}(x,y,w|V)$) and a long run scale term ($CSCE_i\,(x,y,w)$). Notice that $OE_i^{SR}(x,y,w|V)$ = $VC(y,w^v,x^f \mid V)/w^v x^v$ to maintain duality with $DF_i^{SR}(x,y|V,S)$. Since $OE_i(x,y,w \mid V) \lesseqgtr OE_i^{SR}(x,y,w|V)$, clearly $DCU_i^{SR}(x,y,w|V)$ $\lesseqgtr$ 1, while all other terms of the identity are bounded above by unity.

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

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

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

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

$$DCU_i\ (x,y,w|C)\ = OE_i\ (x,y,w|C) / OE_i^{SR}(x,y,w|C).$$

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 $DCU_i\ (x,y,w|C)\ \lesseqgtr 1$ (since $OE_i(x,y,w \mid C) \lesseqgtr OE_i^{SR}(x,y,w|C)$).

This approach allows for some interesting links between the components of these two extended decompositions. Indeed, the short and long run notions of scale efficiency and economic capacity utilisation can be straightforwardly related to one another.[22]

To be more explicit, we discuss the potential integration of the different economic capacity notions in (EDEC1) and (EDEC2) in full detail. First, the first economic capacity notion is easily fitted into (EDEC2), since the minimum of the short run cost function ($OE_i^{SR}(x,y,w|C)$) 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 $DCU_i^{SR}(x,y,w|V)$ implies a tangency point at the long run cost function under *VRS* (i.e., $OE_i(x,y,w \mid V)$). 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:

(EDEC1')
$$OE_i\ (x,y,w|C) = OE_i^{SR}(x,y(p,w^v,x^f),w|V).DCU_i^{SR}(x,y(p,w^v,x^f),y,w|V).CSCE_i\ (x,y,w)$$

where $DCU_i^{SR}(x,y(p,w^v,x^f),y,w|V) = OE_i\ (x,y,w|V) / OE_i^{SR}(x,y(p,w^v,x^f),w|V)$ 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:

(EDEC2')
$$OE_i\ (x,y,w|C) = OE_i^{SR}(x,y(p,w^v,x^f),w|V).CSCE_i^{SR}(x,y(p,w^v,x^f),w).DCU_i^{SR}(x,y(p,w^v,x^f),y,w|C)$$

---

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

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

whereby 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 utilisation notions is made by the scale terms as follows:

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

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

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 $DCU_i^{SR}(x, y(p, w^v, x^f), y, w|C) = OE_i(x, y, w|C) / OE_i^{SR}(x, y(p, w^v, x^f), w|C)$. Notice 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'), since the output level at tangency cost need not correspond to the output level of the evaluated observation, the component measures combining different output levels need not be smaller or equal to unity. Notice furthermore 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, Kerstens and Prior (2009).

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 $CSCE_i(x,y,w)$ respectively $DCU_i^{SR}(x,y,w|C)$, this amounts to re-interpret existing decompositions as measures of capacity utilisation.

To graphically illustrate some of these decompositions, we focus on two concrete definitions integrated into (EDEC1) respectively (EDEC2). First, Figures 3 and 4 illustrate both (EDEC1) and (EDEC2) with the notion of minimal long run average total costs. Next, Figure 5 shows how the tangency cost concept with given fixed inputs can be integrated into (EDEC1). Finally, the minimum of the short run cost function fitting into (EDEC2) is illustrated on the same Figure 4. These examples cover a representative part of the above proposals.

Notice that all Figures 3 to 5 contain the traditional way to measure efficiency along a ray through the origin, but in fact measure technical efficiency components using a short-run radial efficiency measure along the single variable input dimension solely (i.e., along a horizontal line parallel to the axis of the variable input). This serves to integrate the combination of the various technical and economical efficiency components in each of the above decompositions.

Figures 3 and 4 both have $OE_i(x,y,w|C)$ for observation f equal to the ratio $0f_5/0f$. Figure 3 has the following efficiency components: $TE_i^{SR}(x,y) = 0f_2/0f$, $STE_i^{SR}(x,y) = 0f_3/0f_2$, $DCU_i^{SR}(x,y,w|V) = 0f_4/0f_3$, and $CSCE_i(x,y,w) = 0f_5/0f_4$. Note that $AE_i^{SR}(x,y,w) = 1$ for this specific example. In Figure 4, the efficiency components are: $TE_i^{SR}(x,y) = 0f_2/0f$, $STE_i^{SR}(x,y) = 0f_3/0f_2$, $CSCE_i^{SR}(x,y,w) = 0f_4/0f_3$, and $DCU_i^{SR}(x,y,w|C) = 0f_5/0f_4$. Again, $AE_i^{SR}(x,y,w) = 1$. The same Figure 4 also serves to illustrate the case of the minimum short run cost function integrated into (EDEC2). Both capacity notions are specific in that they situate the exact capacity component differently. For minimal long respectively short run average total costs the capacity term is situated in the numerator of $DCU_i^{SR}(x,y,w|C)$ respectively $CSCE_i^{SR}(x,y,w)$. Otherwise, the overall decomposition and its components are identical. Finally, Figure 5 illustrates the tangency cost concept with given fixed inputs and therefore involves a fourth technology with adjusted outputs (relative to the observed outputs of the observation being evaluated). $OE_i(x,y,w \mid C)$ for observation f equals the ratio $0f_4/0f$. Its component measures are: $TE_i^{SR}(x,y(p,w^v,x^f))$ = $0f_2/0f$, $DCU_i^{SR}(x,y(p,w^v,x^f),y,w|V)$ = $0f_3/0f_2$, and $CSCE_i(x,y,w \mid C)$ = $0f_4/0f_3$. Note: $STE_i^{SR}(x,y(p,w^v,x^f))$ = $AE_i^{SR}(x,y(p,w^v,x^f),w)$ = 1.

<FIGURES 3-5 ABOUT HERE>

4.2. 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 utilisation. 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:

(EDEC3) $$OTE_o(x,y) = TE_o(x^f,y).STE_o(x^f,y).PCU_o^{SR}(x,x^f,y|V).SCE_o(x,y),$$

whereby:

$$TE_o(x^f,y) = DF_o(x^f,y \mid V,W);$$

$$STE_o(x^f,y) = DF_o(x^f,y \mid V,S)/DF_o(x^f,y \mid V,W);$$

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

$$SCE_o(x,y) = DF_o(x,y \mid C,S)/DF_o(x,y \mid V,S).$$

Notice 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. Notice 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:

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

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

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

$$PCU_o(x,x^f,y \mid C) = DF_o(x,y \mid C,S)/DF_o(x^f,y \mid C,S).$$

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 utilisation to one another.[23]

5. 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 (see 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 that hydro-power was a dominant source of energy during the 90´s (Pollitt (2004)). Notice 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, Bruvoll and Roar Aune (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. Observe that the minimum price for water is zero, which corresponds to the 11 power plants located on a river (run-of-river plants). For the 5 reservoir plants, the price of water equals the marginal cost of fossil-fueled generation. Table 1 present basic descriptive statistics for the inputs and the single output for the year 1997. More details on the data are available in Atkinson and Dorfman (2009).[24]

[Table 1 around here]

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

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

whereby 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 utilisation notions is provided by the scale terms as follows:

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

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

[24] We maintain all observations rather than opting for a preliminary screening looking for any potential outliers.

Computing the extended decompositions of overall efficiency (EDEC1) to (EDEC4) requires solving a series of optimisation models, since for each observation in the sample all components must be determined using a separate mathematical program. Most of the non-parametric frontier models used in this contribution have already appeared in the literature (see Färe, Grosskopf and Lovell (1994) or Ray (2004)). Therefore, to save some space details on the specifications of the different efficiency measures and cost functions are made available in an Appendix.[25]

Table 2 reports descriptive statistics covering the whole distribution of the efficiency decompositions (EDEC1) to (EDEC4) for the complete sample. 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 33% of observed total costs. In terms of its components, it is clear that a prominent problem comes from the management of the variable inputs since $OE_i^{SR}(x,y,w|V)$ is lowest in both decompositions. Continuing the analysis of the common components in both decompositions, allocative inefficiency ($AE_i^{SR}(x,y,w)$) is the most important problem since it is slightly more acute than technical inefficiency ($TE_i^{SR}(x,y)$). Notice that congestion ($STE_i^{SR}(x,y)$), as a special case of technical efficiency, plays a minor but non-negligible role (about 3%). Shifting focus to the components that differ among decompositions, one notes that scale inefficiencies and capacity utilization show minor differences in magnitude, but lead to the same qualitative conclusions. First, scale inefficiency is the second most important source of poor performance (after $OE_i^{SR}(x,y,w|V)$). Second, capacity utilization is the second largest component after congestion, indicating that errors in the choice of fixed inputs have had on average small cost repercussions. In particular, practically 75% of the sample exhibits a capacity utilization coefficient smaller than unity, implying that the overall inefficiency level relative to total cost is smaller than the one corresponding to variable costs. Since $OE_i^{SR}(x,y,w|V)$ amounts to about 55%, this means that $OE_i(x,y,w|V)$ is even smaller. Thus, the errors made in terms of the choice of fixed inputs are even more costly than the errors made in the management of variable inputs, while the latter are more important than the errors in the choice of scale.

[Table 2 around here]

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

[25] This appendix is available upon simple request.

both decompositions, technical efficiency is very prominent, while congestion is almost negligible (just 0.5%).[26] Turning now to 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). Plant capacity utilization is low, especially in (EDEC4).

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 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 Figure 6, 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 ($DCU_i^{SR}(x,y,w|V)$), 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 ($PCU_o^{SR}(x,x^f,y|V)$), 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.

[Figure 6 around here]

6. CONCLUSIONS

This paper has first reviewed the traditional way of defining different sources of efficiency. Having developed the ways in which both technical and economical capacity utilisation 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

[26] 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.

capacity utilisation. 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.

One possible extension is to derive capacity notions for indirect technologies where output maximizing production is, e.g., subject to a budget constraint (see Ray, Mukherjee and Wu (2006) for non-parametric capacity notions in this context), or for regulated industries (e.g., Segerson and Squires (1993)).

Another 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. Though some first steps have been taken (e.g., Färe, Grosskopf and Kirkley (2000)), discrete time dual productivity indexes could probably equally benefit from the integration of economic capacity terms.[27]

REFERENCES

Atkinson, S.E, J.H. Dorfman, (2009) Feasible Estimation of Firm-Specific Allocative Inefficiency Through Bayesian Numerical Methods, *Journal of Applied Econometrics*, 24(4), 675–697.

Balk, B.M. (2001) Scale Efficiency and Productivity Change, *Journal of Productivity Analysis*, 15(3), 159-183.

Banker, R., A. Charnes, W. Cooper (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30(9), 1078-1092.

Berndt, E., C. Morrison (1981) Capacity Utilization Measures: Underlying Economic Theory and an Alternative Approach, *American Economic Review*, 71(2), 48-52.

Bogetoft, P., R. Färe, B. Obel (2006) Allocative Efficiency of Technically Inefficient Production Units, *European Journal of Operational Research*, 168(2), 450-462.

Bonin, J.P., C.J. Derek, L. Putterman (1993) Theoretical and Empirical Studies of Producer Cooperatives: Will Ever the Twain Meet, *Journal of Economic Literature,* 31(3), 1290-1320.

Briec, W., K. Kerstens, D. Prior (2009) Tangency Capacity Notions Based upon the Profit and Cost Functions: A Non-Parametric Approach and A General Comparison, Lille, IESEG School of Management (Working paper 2009-ECO-05).

Bye, T., A. Bruvoll, F. Roar Aune (2008) Inflow Shortages in Deregulated Power Markets-Reasons for Concern?, *Energy Economics,* 30(4), 1693–1711.

[27] 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).

Cassels, J.M. (1937) Excess Capacity and Monopolistic Competition, *Quarterly Journal of Economics*, 51(3), 426-443.
Caves, R.E. (2007) In praise of the Old I.O., *International Journal of Industrial Organization*, 25(1), 1–12.
Chambers, R., Y. Chung, R. Färe (1998) Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications*, 98(2), 351-364.
Coelli, T., E. Grifell-Tatjé, S. Perelman (2002) Capacity Utilisation and Profitability: A Decomposition of Short-Run Profit Efficiency, *International Journal of Production Economics*, 79(3), 261-278.
De Borger, B., K. Kerstens (2000) The Malmquist Productivity Index and Plant Capacity Utilization, *Scandinavian Journal of Economics*, 102(2), 303-310.
Färe, R., (1984) The Existence of Plant Capacity, *International Economic Review*, 25(1), 209-213.
Färe, R., S. Grosskopf (1985) A Nonparametric Cost Approach to Scale Efficiency, *Scandinavian Journal of Economics*, 87(4), 594-604.
Färe, R., S. Grosskopf (2000) Decomposing Technical Efficiency with Care, *Management Science*, 46(1), 167-168.
Färe, R., S. Grosskopf, J. Kirkley (2000) Multi-Output Capacity Measures and Their Relevance for Productivity, *Bulletin of Economic Research*, 52(2), 101 -112.
Färe, R., S. Grosskopf, E. Kokkelenberg (1989) Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach, *International Economic Review*, 30(3), 655-666.
Färe, R., S. Grosskopf, C.A.K. Lovell (1983) The Structure of Technical Efficiency, *Scandinavian Journal of Economics*, 85(2), 181-190.
Färe, R., S. Grosskopf, C.A.K. Lovell (1985) *The Measurement of Efficiency of Production*, Boston, Kluwer.
Färe, R., S. Grosskopf, C.A.K. Lovell (1994) *Production Frontiers*, Cambridge, Cambridge University Press.
Färe, R., S. Grosskopf, V. Valdmanis (1989) Capacity, Competition and Efficiency in Hospitals: A Nonparametric Approach, *Journal of Productivity Analysis*, 1(2), 123-138.
Farrell, M. (1957) The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120A(3), 253-281.
Førsund, F., L. Hjalmarsson (1974) On the Measurement of Productive Efficiency, *Swedish Journal of Economics*, 76(2), 141-154.
Førsund, F., L. Hjalmarsson (1979) Generalised Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants, *Economic Journal*, 89(354), 294-315.
Fukuyama, H., W. Weber (1999) The Efficiency and Productivity of Japanese Securities Firms, 1988-93, *Japan and the World Economy*, 11(1), 115-133.
Fuss, M., D. McFadden, Y. Mundlak (1978) A Survey of Funtional Forms in the Economic Analysis of Production, in: M. Fuss, D. McFadden (eds) *Production Economics: A Dual Approach to Theory and Applications, Vol. 1*, Amsterdam, North-Holland, 219-268.
Ganley, J., J. Cubbin (1992) *Public Sector Efficiency Measurement: Applications of Data Envelopment Analysis*, Amsterdam, North-Holland.
Ghemawat, P., B. Nalebuff (1985) Exit, *Rand Journal of Economics*, 16(2), 184-194.
Grifell-Tatjé, E., C.A.K. Lovell (2000) Cost and Productivity, *Managerial and Decision Economics*, 21(1), 19-30.
Grosskopf, S. (1986) The Role of the Reference Technology in Measuring Productive Efficiency, *Economic Journal*, 96(382), 499-513.
Hickman, B.G. (1964) On a New Method for Capacity Estimation, *Journal of the American Statical Association*, 59(306), 529-549.

Hulten, C. (1986) Productivity Change, Capacity Utilization, and the Sources of Productivity Growth, *Journal of Econometrics*, 33(1-2), 31-50.

Jamasb, T., M. Pollitt (2001) Benchmarking and Regulation: International Electricity Experience, *Utilities Policy*, 9(3), 107-130.

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, Volume 1, Amsterdam, North Holland, 359-382].

Kallio, M., M. Kallio (2002) Nonparametric Methods for Evaluating Economic Efficiency and Imperfect Competition, *Journal of Productivity Analysis*, 18 (2), 171-189.

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

Kirkley, J., C.J. Morrison Paul, D. Squires (2004) Deterministic and Stochastic Capacity Estimation for Fishery Capacity Reduction, *Marine Resource Economics*, 19(3), 271–294.

Klein, L.R. (1960) Some Theoretical Issues in the Measurement of Capacity, *Econometrica*, 28(2), 272-286.

McDonald, J. (1996) A Problem with the Decomposition of Technical Inefficiency into Scale and Congestion Components, *Management Science*, 42(3), 473-474.

Morrison, C. (1985) Primal and Dual Capacity Utilization: An Application to Productivity Measurement in the U.S. Automobile Industry, *Journal of Business and Economic Statistics*, 3(4), 312-324.

Morrison, C. (1993) *A Microeconomic Approach to the Measurement of Economic Performance: Productivity Growth, Capacity Utilization, and Related Performance Indicators*, New York, Springer Verlag.

Nelson, R. (1989) On the Measurement of Capacity Utilization, *Journal of Industrial Economics*, 37(3), 273-286.

Pascoe, S., D. Tingley (2006) Economic Capacity Estimation in Fisheries: A Non-Parametric Ray Approach, *Resource and Energy Economics*, 28(2), 124-138.

Pollitt, M.G. (2004) Electricity Reform in Chile: Lessons for Developing Countries, *Journal of Network Industries*, 5(3-4), 221–262.

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

Ray, S.C., K. Mukherjee, Y. Wu (2006) Direct and Indirect Measures of Capacity Utilization: A Non-Parametric Analysis of US Manufacturing, *Manchester School*, 74(4), 526-548.

Rowland, W.W., M.R. Langemeier, B.W. Schurle, A.M. Featherstone (1998) A Nonparametric Efficiency Analysis for a Sample of Kansas Swine Operations, *Journal of Agricultural and Applied Economics*, 30(1), 189-199.

Segerson, K., D. Squires (1990) On the Measurement of Economic Capacity Utilization for Multi-Product Industries, *Journal of Econometrics*, 44(3), 347-361.

Segerson, K., D. Squires (1993) Capacity Utilization under Regulatory Constraints, *Review of Economics and Statistics*, 75(1), 76-85.

Segerson, K., D. Squires (1995) Measurement of Capacity Utilization for Revenue-Maximizing Firms, *Bulletin of Economic Research*, 47(1), 77-85.

Seitz, W.D. (1970) The Measurement of Efficiency Relative to a Frontier Production Function, *American Journal of Agricultural Economics*, 52(4), 505-511.

Seitz, W.D. (1971) Productive Efficiency in the Steam-Electric Generating Industry, *Journal of Political Economy*, 79(4), 878-886.

Squires, D. (1987) Long-Run Profit Functions for Multiproduct Firms, *American Journal of Agricultural Economics*, 69(3), 558-569.

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(11), 1593-1608.

Table 1. Descriptive Statistics for 1997

Variable	Trimmed Mean†	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 (# 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 labor (millions per worker)	1.26	1.23	1.28
Price of capital (estimated cost of capital)	0.70	0.63	0.77

† 10% trimming level.

Table 2. Descriptive Statistics for the Efficiency Decompositions (EDEC1-EDEC4)

		Trimmed Mean†	10th Percentile	1st Quartile	Median	3rd Quartile	90th Percentile
EDEC1	$OE_i\ (x,y,w\|C)$	0.3292	0.0562	0.1313	0.2656	0.5511	0.6589
	$OE_i^{SR}(x,y,w\|V)$	0.5498	0.1429	0.2853	0.5707	0.8289	0.9984
	$DCU_i^{SR}(x,y,w\|V)$	0.9287	0.7323	0.8680	0.9972	1.0073	1.0150
	$CSCE_i\ (x,y,w)$	0.7359	0.3292	0.5741	0.7992	0.9140	0.9590
	$TE_i^{SR}(x,y)$	0.7739	0.3402	0.5862	0.8299	0.9994	1.0000
	$STE_i^{SR}(x,y)$	0.9733	0.8367	0.9847	0.9994	1.0000	1.0000
	$AE_i^{SR}(x,y,w)$	0.7684	0.1912	0.5734	0.8266	0.9911	1.0000
EDEC2	$OE_i\ (x,y,w\|C)$	0.3292	0.0562	0.1313	0.2656	0.5511	0.6589
	$OE_i^{SR}(x,y,w\|V)$	0.5498	0.1429	0.2853	0.5707	0.8289	0.9984
	$DCU_i\ (x,y,w\|C)$	0.8989	0.8195	0.8526	0.8957	0.9832	1.0754
	$CSCE_i^{SR}(x,y,w)$	0.7406	0.3593	0.5302	0.8338	0.9220	0.9692
	$TE_i^{SR}(x,y)$	0.7739	0.3402	0.5862	0.8299	0.9994	1.0000
	$STE_i^{SR}(x,y)$	0.9733	0.8367	0.9847	0.9994	1.0000	1.0000
	$AE_i^{SR}(x,y,w)$	0.7684	0.1912	0.5734	0.8266	0.9911	1.0000
EDEC3	$OTE_o(x,y)$	1.5753	1.0057	1.1396	1.4294	1.9752	2.5565
	$TE_o(x^f,y)$	3.7302	1.2009	1.8071	2.5880	5.0129	9.1133
	$STE_o(x^f,y)$	1.0057	1.0000	1.0000	1.0000	1.0000	1.0416
	$PCU_o^{SR}(x,x^f,y\|V)$	0.5226	0.1472	0.3354	0.5334	0.7188	0.8768
	$SCE_o(x,y)$	1.0953	1.0000	1.0108	1.0376	1.1824	1.2863
EDEC4	$OTE_o(x,y)$	1.5753	1.0057	1.1396	1.4294	1.9752	2.5565
	$TE_o(x^f,y)$	3.7302	1.2009	1.8071	2.5880	5.0129	9.1133
	$STE_o(x^f,y)$	1.0057	1.0000	1.0000	1.0000	1.0000	1.0416
	$PCU_o(x,x^f,y\|C)$	0.3814	0.0826	0.2065	0.3923	0.5076	0.7288
	$SCE_o^{SR}(x^f,y)$	1.5695	1.0085	1.2499	1.6218	1.7797	2.1927

† 10% trimming level.

Notice that to facilitate comparisons in the table the order of the components of (EDEC2) and (EDEC4) follows the order of the decompositions (EDEC1) and (EDEC3).

Figure 1: DEC1 & DEC2 Illustrated on Input Sets with Different Production Axioms

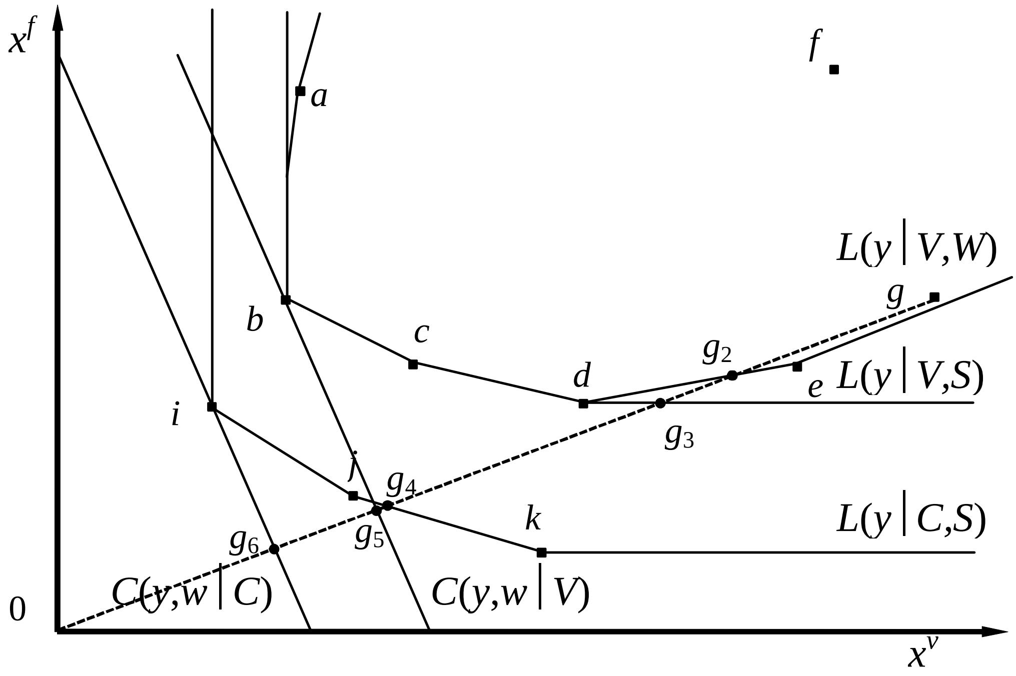

Figure 2: Different Notions of Cost-based Capacity Utilisation

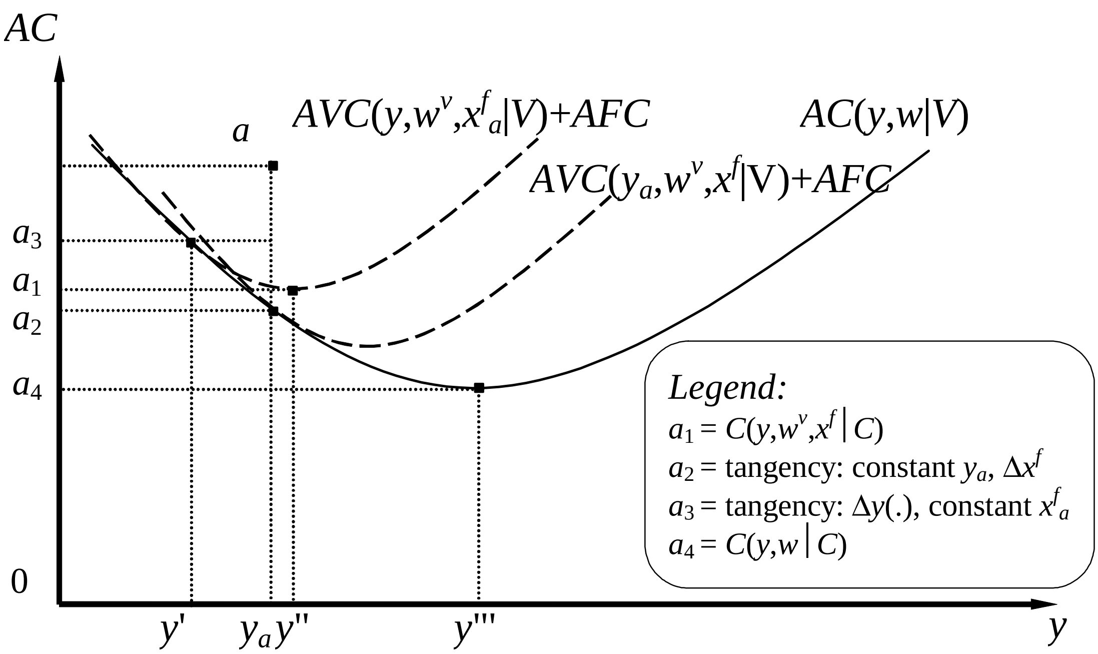

Figure 3: EDEC1 with Minimum of the Long Run Average Total Cost Function

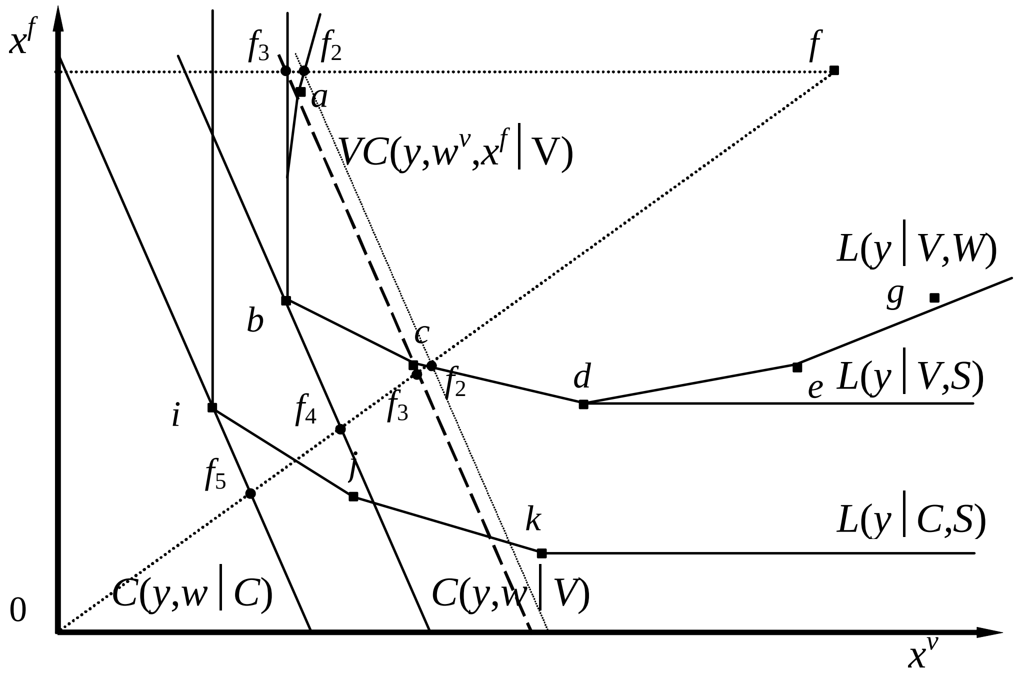

Figure 4: EDEC2 with Minimum of the Short and Long Run Average Total Cost Function

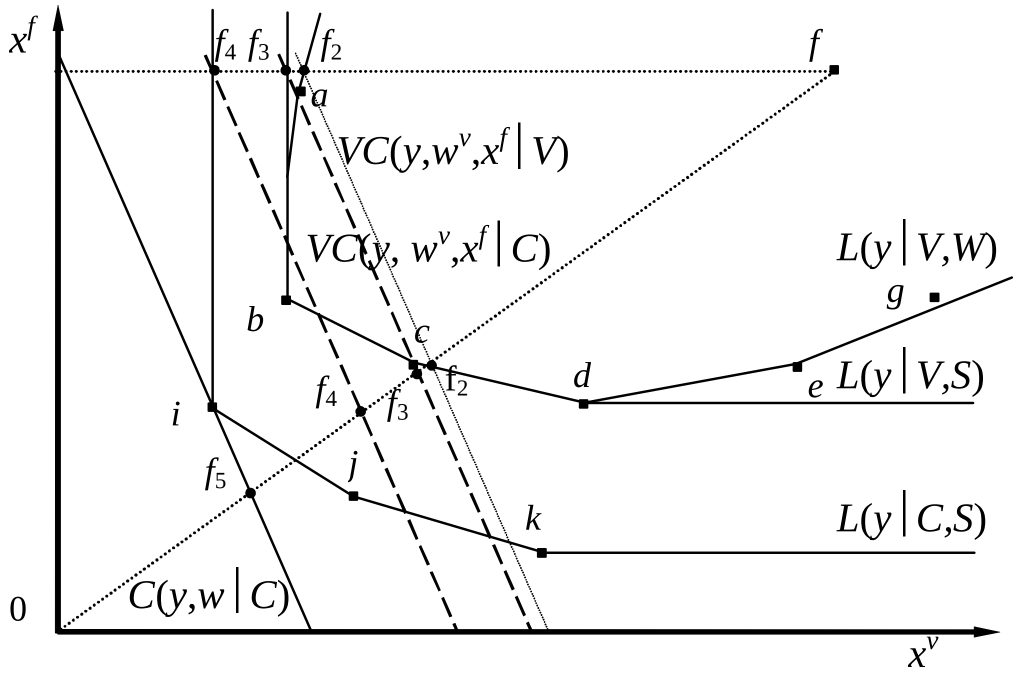

Figure 5: EDEC1 with Tangency Cost Notion (x^f fixed; y adjusted)

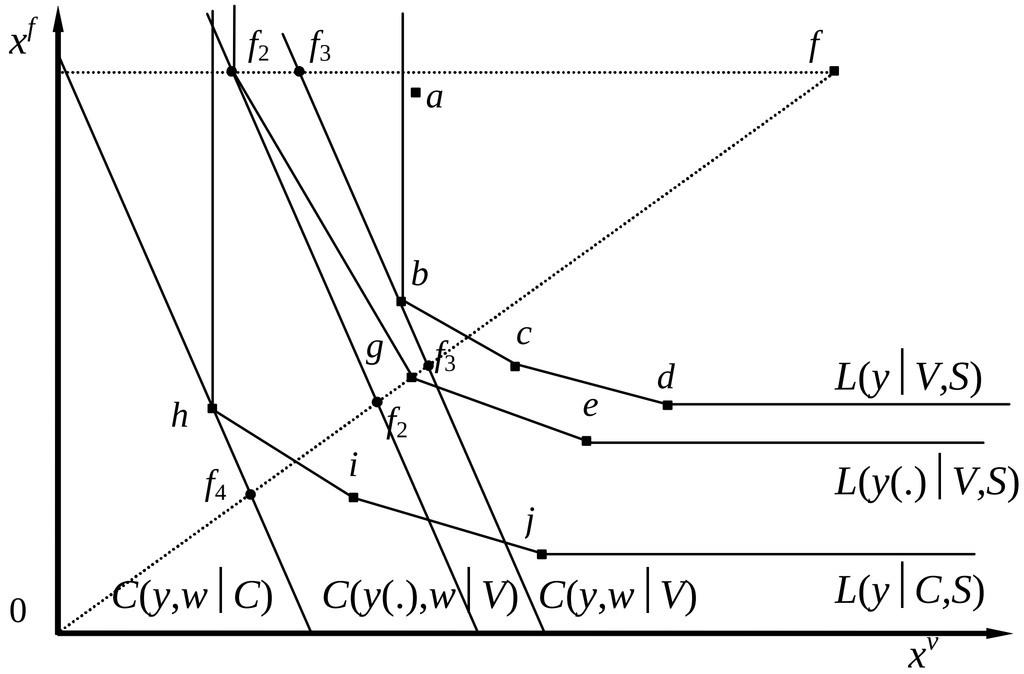

Figure 6: Average Monthly Capacity Components for Run-of-River vs. Reservoir Plants (DCU (EDEC1) and PCU (EDEC4))