Industry structural inefficiency and potential gains from mergers and break-ups: an empirical approach

Antonio Peyrache

Centre for Efficiency and Productivity Analysis (CEPA)
School of Economics, The University of Queensland, Australia
(a.peyrache@uq.edu.au)

Abstract

In this paper an encompassing empirical strategy is presented which is able to decompose an indicator of industrial structural inefficiency into its sources components. The main purpose of the analysis is to bring a number of ideas floating in the efficiency literature together in the same empirical industrial organization model and extend them to a more comprehensive definition of industry inefficiency. The tool used to reach this goal is the directional distance function (DDF) representation of the data envelopment analysis (DEA) data generated technology. As it is standard in the linear activity analysis model, all the discussion is based only on the assumption that measures of inputs and outputs quantities are available, without reference to prices. Decomposing the industrial structural inefficiency indicator into different components the following effects have been identified: 1) inefficiencies arising from firms operating on a large size that can be split into smaller more productive units (size inefficiencies); 2) efficiency gains that can be realized thanks to the merger of small firms (merger inefficiencies); 3) re-allocation of inputs and outputs in order to bring firms toward an optimal production plan (re-allocation inefficiencies). After defining the static industry inefficiency indicator, a dynamic decomposition of productivity change will be proposed. Productivity change itself is decomposed into technical change and efficiency components. The methodology is applied to healthcare data on public hospitals in Australia. The empirical results point to the fact that technical inefficiency of individual hospitals accounts only for less than 15% of the total inefficiency of the industry. The most part of industry inefficiency has been found to be organizational. Size inefficiency is the most prominent component accounting for around 40% of the total inefficiency of the industry.

Key words: industry inefficiency; size inefficiency; mergers; break-ups; shadow profit; directional distance function

Jel code: D24, L23, M11

1. Introduction

The measurement of industrial efficiency has important implications in practical terms. As pointed out by the industrial organization literature, monopolies, break-ups of monopolies and mergers of firms are all justified or opposed on the ground that they can produce (at least in the long run) lower production costs (which is the dual to an increase in production efficiency) and increased welfare. Information on the potential gains from these operations is particularly important when antitrust authorities and regulators have to approve a merger or decide on the break-up of a large firm. Things are even more complicated in the public sector where production units respond to non-market incentives and these incentives can lead to strongly sub-optimal configurations of the industry. The key point here is that in many public sector industries (such as health and education) it is not possible to determine easily output prices, even when output quantities and costs can be somehow more precisely defined. In both the market and the non-market scenarios the main question relies on what is an efficient configuration of the industry and how we can measure deviations from this optimal configuration and attribute it to different sources.

The empirical strategy here introduced can find application also as an operations research tool. Mutatis mutandis, the previous arguments can also be applied to the management of large companies which are operating with a number of different production plants. In this last case the phrasing is different but the essence is the same: the profit of the company crucially depends on the way its production is allocated across the different production plants. In this interpretation the question of which configuration of the company is efficient is different in nature from the industry configuration problem but it relies on the same method. Therefore the proposed framework can be used to organize efficiently a single large company by assigning its total production to an optimal number of production units. This has implication in the restructuring of large companies where estimation of potential gains from re-organization is needed and monitoring of the actual productivity increase deriving from the re-organization must be assessed. For the sake of simplicity the industrial organization phrasing (industry vs firms) will be used, though at any stage of the argument one can interpret the industry as a single large company and the firms as its production unit components.

The empirical strategy presented in this paper builds on previous results stemming from the efficiency literature. The method is based only on the knowledge of input and output quantities (no price information is required), therefore applicable both to market and non-market oriented industries. Since the seminal work of Banker et al (1984) data envelopment analysis (DEA) has been used to provide information on the technical efficiency with which individual firms in an industry are operating. This has also been extended to a more dynamic framework of productivity measurement more recently (see for example Banker et al, 2005). Many of those contributions did not distinguish explicitly between the firm and the industry level. In fact, an industry can be inefficiently organized even if all the individual firms which compose it operate efficiently, i.e. efficiency of the firms is a necessary but not sufficient condition for the efficiency of the industry. Thus the first task of this paper is to define a general measure of industry efficiency as distinguished from the efficiency of the firm. Three main research trajectories are relevant to the results here developed. *First*, there are contributions trying to identify if an efficiency gain can be obtained at industry level by splitting large firms into an optimal number of smaller firms: this relates to the notion of size efficiency presented by Maindiratta (1990), Ray (2007) and Ray and Mukherjee (1998). This approach was also followed by Newbery and Pollitt (2007) on their discussion of the restructuring of the electricity sector in the UK. *Second*, it is possible to identify a number of contributions trying to understand when a merger brings efficiency gains. Here there are two possible causes of this increase in efficiency: on the one hand the merged firms can have stronger incentives towards technical efficiency (see Banal-Estanol, 2008, Weber and Camerer, 2003 for a theoretical discussion and Kwoka and Pollitt, 2010 for an empirical investigation); on the other hand the merged firms can benefit from scale and scope economies once merged (this is the way followed by Bogetoft et al, 2003, Bogetoft and Wand, 2005 and Kristensen et al, 2010). *Third*, there are two contributions which go in the direction of defining an overall measure of performance for the industry: Briec et al (2003) define an index of industry efficiency based on the assumption that the number of firms in the industry is fixed; Ray and Hu (1997) define a similar index allowing the number of firms to be variable.

The main purpose of this study is to bring these ideas together in the same empirical model and extend them to a more comprehensive definition of industry inefficiency and its components. The tool used to reach this goal is the directional distance function (DDF) representation of the production technology. The extension of all the previous notions to a multi-period analysis of productivity measurement is introduced. An empirical application on health data (hospital level observations) is presented and it is shown that technical efficiency accounts for as low as 15% of total industry efficiency, with around 85% of this inefficiency attributable to the way the industry is organized.

The rest of the paper is organized as follow. Section 2 introduces the technology and its functional description via the directional distance function. Section 3 defines industry inefficiency and its components. Section 4 extends the analysis to a dynamic context in which the industry is compared in different time periods. Section 5 introduces an index of overall firm inefficiency based on shadow pricing of inputs and outputs. Section 6 presents the empirical application and, finally, section 7 concludes.

2. Technology

Consider an industry where $\boldsymbol{x} \in \mathbb{R}_+^N$ inputs produce $\boldsymbol{y} \in \mathbb{R}_+^M$ outputs. We assume an unbalanced panel data specification: in each time period data on inputs and outputs $(\boldsymbol{x}_k, \boldsymbol{y}_k)$ are available for a number K_t of firms ($k=1,...,K_t$); the total number of time periods is T and for each time period observations on inputs can be collected in a K_txN matrix X_t and observations on outputs can be collected in a K_txM matrix Y_t:

$$(\boldsymbol{X}_t, \boldsymbol{Y}_t), \quad t = 1, \dots, T \tag{1}$$

This specification is general enough to allow interpreting the industry as a single large firm in which K_t separate production units operate in each time period. The production possibility set at time t is the set of all feasible combinations of inputs and outputs producible with the technology and can be approximated by data generated technologies. The first technology to be discussed involves assuming the free disposability axiom (axioms are independent of time and assumed to hold in each time period):

A1) Free disposability: $\forall(\boldsymbol{x},\boldsymbol{y}) \in T, \boldsymbol{x}' \geq \boldsymbol{x}\ and\ \boldsymbol{y}' \leq \boldsymbol{y} \Rightarrow (\boldsymbol{x}',\boldsymbol{y}') \in T$

The associated technology is known as the free disposal hull (FDH) technology and relies only on axiom A1. In order to give more structure to the technology the following axiom will also be considered:

A2) Convexity: $\forall(\boldsymbol{x},\boldsymbol{y}) \in T, (\boldsymbol{x}',\boldsymbol{y}') \in T\ and\ 0 \leq \theta \leq 1 \Rightarrow (\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}', \theta\boldsymbol{y} + (1-\theta)\boldsymbol{y}') \in T$

Based on these two axioms (A1 and A2) the data envelopment analysis (DEA) data generated technology at time t is given by the following definition:

$$T^t = \{(\boldsymbol{x},\boldsymbol{y}): \boldsymbol{\lambda X^t} \leq \boldsymbol{x}, \boldsymbol{\lambda Y^t} \geq \boldsymbol{y}, \boldsymbol{1}_K\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \boldsymbol{0}\} \tag{2}$$

This is a variable returns to scale technology where the intensity vector $\boldsymbol{\lambda}$ is constrained to sum up to one. A complete functional representation of this production possibility set is given by the directional distance function (DDF) (see Chambers et al, 1996, 1998):

$$D^t(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y) = \sup_\beta\{\beta > 0: (\boldsymbol{x} - \boldsymbol{g}_x\beta, \boldsymbol{y} + \boldsymbol{g}_y\beta) \in T^t\} \tag{3}$$

The DDF is searching the maximum expansion of outputs and contraction of inputs along the direction $(\boldsymbol{g}_x,\boldsymbol{g}_y) \geq \boldsymbol{0}$ which is feasible with technology (2). The DDF can be interpreted as a measure of absolute technical inefficiency, representing the physical output loss and input waste of the firm measured in terms of the numeraire $(\boldsymbol{g}_x,\boldsymbol{g}_y)$. In other words equation (3) is a measure of deviation of the specific production plan used by the firm from the benchmark production frontier: any deviation from the production frontier will result in a strictly positive value of function (3). Problem (3) is a standard linear program and can be easily solved once data are available. An additional axiom which will come useful in the following analysis is (Fare, 1986):

A3) Additivity: $\forall\ (\boldsymbol{x}_1,\boldsymbol{y}_1) \in T^t\ and\ (\boldsymbol{x}_2,\boldsymbol{y}_2) \in T^t, (\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}_1 + \boldsymbol{y}_2) \in T^t$

Under A1, A2 and A3 the DEA production technology becomes additive and can be written as:

$$T_A^t = \{(\boldsymbol{x},\boldsymbol{y}): \boldsymbol{\lambda X} \leq \boldsymbol{x}, \boldsymbol{\lambda Y} \geq \boldsymbol{y}, \boldsymbol{1}_K\boldsymbol{\lambda} = S, \boldsymbol{\lambda} \geq \boldsymbol{0}, S \in \mathbb{N}\} \tag{4}$$

In this technology the VRS constraint has been replaced by a generic natural number $S \in \mathbb{N}$: this guarantees additivity of the technology. The subscript 'A' serves to emphasize that the additive technology is an enlargement of the actual VRS technology: $T_A^t \supseteq T^t$. It is possible to associate an additive DDF as a description of this enlarged additive technology (4):

$$D_A^t(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y) = sup_\beta\{\beta > 0: (\boldsymbol{x} - \boldsymbol{g}_x\beta, \boldsymbol{y} + \boldsymbol{g}_y\beta) \in T_A^t\} \tag{5}$$

This last definition of DDF implies solving a mixed integer linear program. The optimal solution will return the intensity vector $\boldsymbol{\lambda}$, the value of the DDF β and the optimal level S_t^* for the intensity vector constraint. This integer value S_t^* can be interpreted (as will be shown more clearly in the next section) as the optimal number of production plants who would produce a given output efficiently. Since the VRS technology is a subset of the additive technology ($T_A^t \supseteq T^t$), the following relationship holds in terms of the associated DDFs: $D_A^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y) \geq D^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y)$. An alternative enlargement of the VRS technology is given considering the following axiom:

A4) Divisibility: $\forall\ (\boldsymbol{x}, \boldsymbol{y}) \in T^t\ and\ 0 \leq \mu \leq 1, (\mu\boldsymbol{x}, \mu\boldsymbol{y}) \in T^t$

Under A1, A2 and A4 the DEA non-increasing returns to scale technology is considered:

$$T_{NIRS}^t = \{(\boldsymbol{x}, \boldsymbol{y}): \boldsymbol{\lambda X} \leq \boldsymbol{x}, \boldsymbol{\lambda Y} \geq \boldsymbol{y}, \mathbf{1}_K\boldsymbol{\lambda} \leq \mathbf{1}, \boldsymbol{\lambda} \geq \mathbf{0}\} \quad (6)$$

Associated to the NIRS technology the following DDF is defined:

$$D_{NIRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y) = \sup_\beta\{\beta > 0: (\boldsymbol{x} - \boldsymbol{g}_x\beta, \boldsymbol{y} + \boldsymbol{g}_y\beta) \in T_{NIRS}^t\} \quad (7)$$

Since $T_{NIRS}^t \supseteq T^t$ then $D_{NIRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y) \geq D^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y)$. There is no simple relationship between the NIRS and the additive technology and therefore no simple relation between the associated directional distance functions. If the last two axioms are considered jointly a further enlargement of the technology set is considered:

A5) Constant returns to scale (CRS): $\forall\ (\boldsymbol{x}, \boldsymbol{y}) \in T^t\ and\ \mu \geq 0, (\mu\boldsymbol{x}, \mu\boldsymbol{y}) \in T^t$

It is important to note that axiom A5 is not a primitive axiom: it is the results of the joint satisfaction of axioms A3 and A4. In fact, A5 holds if and only if axioms A3 and A4 hold. Using A5 jointly with A1 and A2 (i.e. using A1, A2, A3 and A4) the following DEA technology is obtained:

$$T_{CRS}^t = \{(\boldsymbol{x}, \boldsymbol{y}): \boldsymbol{\lambda X} \leq \boldsymbol{x}, \boldsymbol{\lambda Y} \geq \boldsymbol{y}, \boldsymbol{\lambda} \geq \mathbf{0}\} \quad (8)$$

The CRS technology is an enlargement of all the previous technologies and can be described in a functional form by the following DDF:

$$D_{CRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_x, \boldsymbol{g}_y) = sup_\beta\{\beta > 0: (\boldsymbol{x} - \boldsymbol{g}_x\beta, \boldsymbol{y} + \boldsymbol{g}_y\beta) \in T_{CRS}^t\} \quad (9)$$

Since $T_{CRS}^t \supseteq T_A^t \supseteq T^t$ and $T_{CRS}^t \supseteq T_{NIRS}^t \supseteq T^t$, the following relationships among the four different directional distance functions hold:

$$D_{CRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y) \geq D_A^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y) \geq D^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y)$$

$$D_{CRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y) \geq D_{NIRS}^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y) \geq D^t(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y)$$

In other words the CRS-DDF is an upper bound and the VRS-DDF is a lower bound, with the other two functions being in the middle. All the previous DDFs satisfy the following property (see Fukuyama, 2003 Chambers et al, 1998 for a full discussion of DDF properties):

P1) $D^t(\boldsymbol{x},\boldsymbol{y};\alpha\boldsymbol{g}_x,\alpha\boldsymbol{g}_y)=\frac{1}{\alpha}D^t(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y), \alpha>0$

The importance of property P1 relies on the fact that the level of the directional vector is not of a substantive nature: a re-scaling of the directional vector can be associated to an inflated DDF value. Moreover the additive and CRS based DDFs also satisfy the following properties:

P2) $D_A^t(S\boldsymbol{x},S\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y)=S\cdot D_A^t(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y), S\in\mathbb{N}$

P3) $D_{CRS}^t(\mu\boldsymbol{x},\mu\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y)=\mu D_{CRS}^t(\boldsymbol{x},\boldsymbol{y};\boldsymbol{g}_x,\boldsymbol{g}_y), \mu\in\mathbb{R}^+$

It should be emphasized that the additive, NIRS and CRS technologies are enlargements of the basic VRS technology. In what follows the VRS technology is considered the real technology that firms faces and the additive, NIRS and CRS technologies as tools for investigating the structure of the industry (i.e., not all the points in the additive, NIRS and CRS are really feasible production plans for a single firm).

3. Static: Industry vs firm inefficiency

Since the DDF (3) is a measure of firm technical inefficiency, it is possible to define an industry aggregate technical inefficiency indicator (ITE) as the sum of individual firm technical inefficiencies as measured by the DDF:

$$ITE^t=\sum_{k=1}^{K_t}D^t(\boldsymbol{x}_k^t,\boldsymbol{y}_k^t;\boldsymbol{g}_x,\boldsymbol{g}_y) \qquad (10)$$

This summation operation is allowed due to the additive nature of the DDF (inefficiency is expressed in terms of a given numeraire $(\boldsymbol{g}_x,\boldsymbol{g}_y)$). This measure of aggregate inefficiency represents the loss in total industry outputs and waste in total industry inputs due to technical inefficiencies of individual firms composing the industry (firms who do not produce onto the benchmark production frontier). The observed industry total inputs and outputs are: $\boldsymbol{I}^t=\sum\boldsymbol{x}_k^t$, $\boldsymbol{Q}^t=\sum\boldsymbol{y}_k^t$. From the perspective of the overall industry, technical inefficiency is only one source of inefficiency, the other one having to do with the organization of the industry itself. An overall measure of industry structural efficiency (IE) is given by the following mixed integer linear program:

$$IE^t=D_A^t(\boldsymbol{I}^t,\boldsymbol{Q}^t;\boldsymbol{g}_x,\boldsymbol{g}_y) \qquad (11)$$

This indicator of industry efficiency considers the total inputs and outputs of the industry as given and is trying to benchmark them against an optimal value given by the additive technology: it is equal to zero if the industry is operating efficiently and larger than zero if the industry is operating inefficiently. The additive technology is chosen as a benchmark because, from an industry perspective, the number of firms is a variable and should be chosen in order to squeeze the maximum output from the available inputs. As a by-product of this mixed linear program one also determines the optimal number of firms S_t^* which would operate the industry most efficiently. Thus, the mixed integer problem (11) returns simultaneously the optimal number of firms that should populate the industry and the total loss in output and waste in input of the industry. Using property P2 of the additive DDF the industry inefficiency indicator (11) can be written as:

$$IE^t = S_t^* D^t\left(\bar{\boldsymbol{I}}^{\boldsymbol{t}}, \bar{\boldsymbol{Q}}^{\boldsymbol{t}}; \boldsymbol{g}_x, \boldsymbol{g}_y\right) \tag{12}$$

where $(\bar{\boldsymbol{I}}^{\boldsymbol{t}}, \bar{\boldsymbol{Q}}^{\boldsymbol{t}}) = \frac{1}{S_t^*}(\boldsymbol{I}^t, \boldsymbol{Q}^t)$ is a scaling of the vector of sample average of inputs and outputs. In other words industry inefficiency can also be expressed as a multiple of the inefficiency of a hypothetical firm producing the sample average industry outputs using the sample average inputs and benchmarked against the actual VRS technology. The point $(\bar{\boldsymbol{I}}^{\boldsymbol{t}}, \bar{\boldsymbol{Q}}^{\boldsymbol{t}})$ is particularly important because represents the production plan which would maximize industry efficiency if adopted by all the firms of the industry. In fact, Maindiratta (1990) showed that any optimal configuration of the industry in which firms use different input and output bundles is equivalent to a configuration in which S_t^* identical firms produce the same input-output vectors. This optimal vector is $\left(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t\right) = \frac{1}{S_t^*}\left(\boldsymbol{I}^{\boldsymbol{t}} - \boldsymbol{g}_x IE^t, \boldsymbol{Q}^{\boldsymbol{t}} + \boldsymbol{g}_y IE^t\right)$ and is sitting onto the VRS technology. Thus the overall industry inefficiency is a multiple of the inefficiency of a specific optimal point belonging to the VRS technology and the purpose of the mixed integer linear program (12) is basically to identify the optimal number of firms that should populate the industry. Therefore an industry configuration in which S_t^* identical firms produce according to the production plan $\left(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t\right)$ will return an industry which is producing efficiently. This optimal configuration is the benchmark for the industry, though it is possible, in principle, to find other configurations which give rise to the same level of efficiency.

Since an industry can be organized inefficiently even if all the individual firms which compose it operate efficiently onto their production frontier, the following relation between the previous two indicators is established $IE^t \geq ITE^t$ and the discrepancy between the two indicators gives rise to an industry organizational inefficiency indicator (IOE):

$$IOE^t = IE^t - ITE^t \tag{13}$$

Fare et al (2008) showed a result similar to (13) where the number of firms is fixed. An important feature of indicator (13) is that it is defined directly in aggregate terms, stressing the fact that the industry perspective cannot be fully recovered only by the knowledge of the firms composing it. Our technology allows for variable returns to scale and scope economies (i.e., it is non-additive), therefore the indicator of organizational inefficiency (13) accounts for three very different sources of inefficiency: *first*, there are firms which are too large and the industry would benefit from splitting these big firms into smaller ones producing on a more productive scale; *second*, there are firms which are producing on a small scale and could benefit from mergers with other small firms, both in terms of scale and scope (synergies in production of different outputs); *third*, there could be reallocations of inputs and outputs among firms which lead to a gain in total industry production. As a point of clarification it must be said that this study is not searching for aggregation across firms as done in the works of Fare and Primont (2003), Fare and Zelenyuk (2002, 2003) and Zelenyuk (2006). It is rather defining an aggregate industry efficiency indicator and then trying to decompose it into meaningful components.

2.1. Break-ups and size efficiency

Break-up of a single large firm can be investigated using the notion of size efficiency introduced by Maindiratta (1990) and further developed in Ray and Mukherjee (1998) in a traditional distance function formulation. Size efficiency is formally defined as the gap between the additive and the VRS technology at a specific point and can be defined in terms of the DDF as:

$$SE^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y) = D_A^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y) \tag{14}$$

If $SE^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y) > 0$ then a gain in production may be obtained at the industry level by splitting the firm into different units, with the optimal number of units given by the solution to the mixed integer linear program $D_A^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y)$. Due to the additive nature of the DDF, the individual firm size inefficiencies can be summed up into a measure of aggregate industry size efficiency (ISE):

$$ISE^t = \sum_{k=1}^{K_t} SE^t(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y) \tag{15}$$

This last indicator represents the gain in production that could be obtained if all the large firms were to be split into an optimal number of smaller units. It should be stressed that the efficiency gain is at the industry level because it can be well the case that the objective of the firm is to produce on a large scale: in this instance the objective of the firm and the objective of the industry could be conflicting.

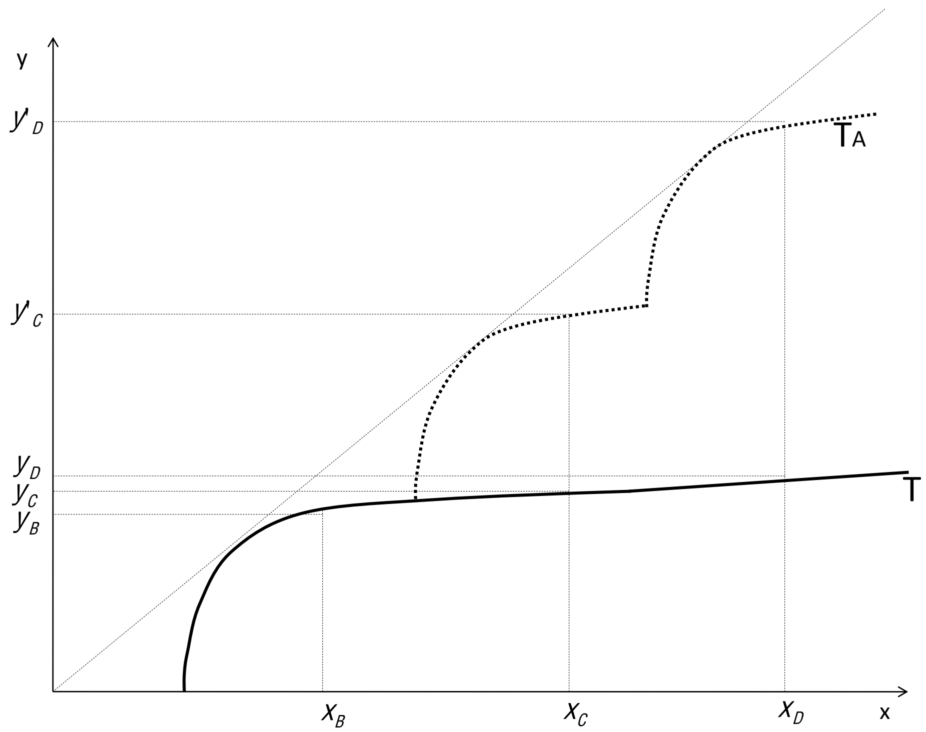


Figure 1 – Example of a one input-one output technology

To provide a better intuition a diagrammatic representation is provided in figure 1. The industry technology T and the enlarged additive technology T_A are shown in the figure. Firm B is size efficient because no gain in production can be obtained at the industry level by splitting this firm into two separate firms. It should be noted that firm B is operating in a decreasing returns to scale region, which means decreasing returns are not sufficient for size inefficiency to arise. On the contrary firm C is size inefficient; in fact by splitting it into two separate firms a gain in production (at industry level) of $SE = y'_C - y_C$ can be obtained. Firm D is size inefficient as well and industry output can be increased by splitting this firm into 3 separate firms with a gain in production of $SE = y'_D - y_D$. In this example the industry composed by the 3 firms can obtain a total gain in production by allocating the same total amount of inputs among 6 different size efficient firms.

3.2. Mergers

A merger between firm A and firm B is efficient if (see Bogetoft and Wand, 2005 for a traditional distance function formulation):

$$ME^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) = D^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D^t(\boldsymbol{x}_A^t, \boldsymbol{y}_A^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D^t(\boldsymbol{x}_B^t, \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) \geq 0 \quad (16)$$

The combined firm has more potential for expansion of outputs and contraction of inputs than the two firms taken separately. It is interesting to note that this corresponds to the definition of a (locally) super-additive technology. An additional condition needed in order for the merger to be fully efficient is that the resulting merged firm is size efficient:

$$SE^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) = 0 \tag{17}$$

Condition (17) is needed because otherwise the merged firm could be split into two separate units obtaining a gain in production. It could be that condition (16) holds (the technology is locally super-additive) but condition (17) does not hold (the merged firm can be split with a gain in production). This last case is not considered a merger, but a simple re-allocation of inputs and outputs between the two original firms, in which case:

$$D^t(\boldsymbol{x}_A^t, \boldsymbol{y}_A^t; \boldsymbol{g}_x, \boldsymbol{g}_y) + D^t(\boldsymbol{x}_B^t, \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) \le 2D^t\left(\frac{\boldsymbol{x}_A^t + \boldsymbol{x}_B^t}{2}, \frac{\boldsymbol{y}_A^t + \boldsymbol{y}_B^t}{2}; \boldsymbol{g}_x, \boldsymbol{g}_y\right)$$

This last formula point to the fact that the potential efficiency gain of the average firm is larger than the gain obtainable by the two original firms and in order to reach this gain a re-allocation of inputs and outputs between the two firms must be pursued to push them towards their mean.

In order to assess if a merger is efficient formulas (16) and (17) are used: formula (16) will check that the technology is locally super-additive and formula (17) will check that the resulting merged firm is size efficient. A tree algorithm for the identification of all possible mergers is as follow:

1. Determine all size inefficient firms and split them in the optimal number of sub-units, obtaining a new dataset with $J \ge K$ number of firms.
2. Consider all the possible combinations without repetition of the J firms taken 2 at time;
3. Compute the following vector of indicators in which each element is computed for each pair of the $\frac{J(J-1)}{2}$ combinations:
   a) $ME^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y)$ if $SE^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) = 0$
   b) 0 if $SE^t(\boldsymbol{x}_A^t + \boldsymbol{x}_B^t, \boldsymbol{y}_A^t + \boldsymbol{y}_B^t; \boldsymbol{g}_x, \boldsymbol{g}_y) > 0$
4. Choose the largest value of the vector and merge the two associated firms obtaining J-1 firms and the indicator of gain in efficiency from the merger ME_j^t. If the largest value of the vector is equal to zero, then there are no mergers that can lead to gain in efficiency, otherwise repeat from 2) with J-1 firms.

Summing up all the previous indexes from the mergers returns an industry measure of merger inefficiency (IME^t): this index is positive if some gain in efficiency can be obtained at industry level by merging two firms and is equal to zero if no gains in efficiency can arise from mergers.

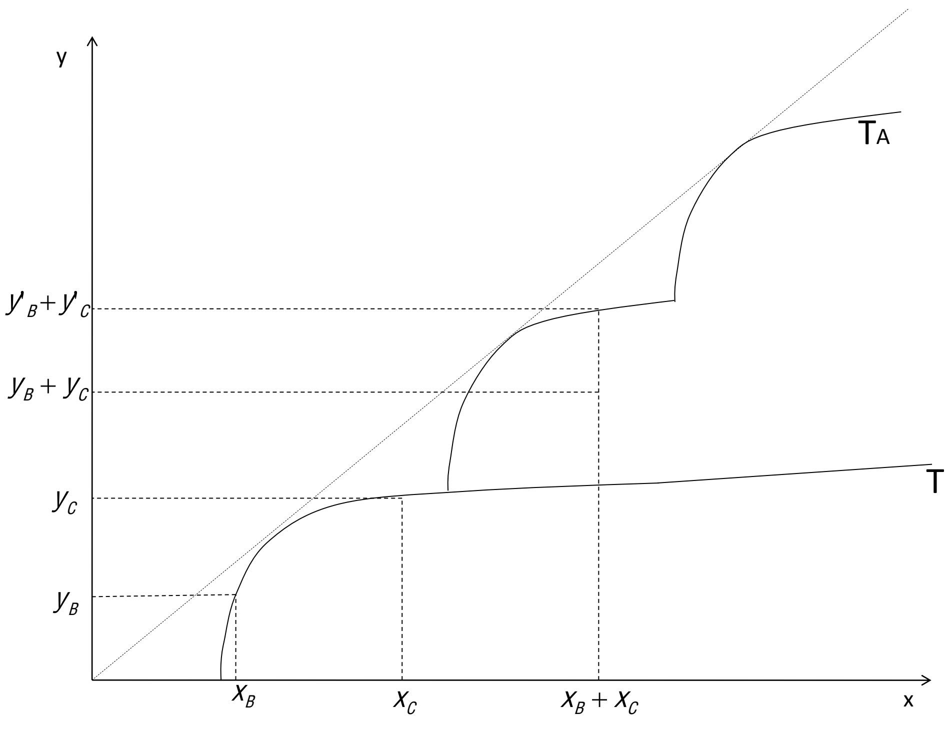
3.3. Re-allocation of inputs and outputs

After the previous discussion it is possible to think of the industry as structured in a configuration where there are no possible mergers or break-ups that can lead to gains in efficiency. Even with such a configuration, the industry organizational inefficiency indicator (13) can be positive due to possible reallocations of inputs and outputs among firms. Thus an index of industry reallocation inefficiency (RE) can be defined in aggregate terms in a residual way as:

$$IRE^t = IOE^t - IME^t - ISE^t \tag{18}$$

In fact, any configuration of the industry which deviates from S_t^* identical firms producing $\left(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t\right)$ can give at the most the same level of efficiency as the S_t^* identical firms configuration. Thus, at the end of break-ups and mergers there can be some space for further improvements thanks to reallocations of inputs and outputs among firms.

The notion of gain in efficiency due to reallocation can be easily illustrated by the following example of an industry composed by 2 firms. The total amount of industry input $x_B + x_C$ can produce an output of $y'_B + y'_C$ if input is allocated efficiently among firms. On the contrary with the given allocation the industry is producing a sub-optimal output of $y_B + y_C$. This means that reallocating inputs from firm C to firm B will increase industry output up to the efficient level.



3.4. Structural decomposition of industry efficiency

With the previous explanation of single components which induce inefficiency at the aggregate industry level, the overall decomposition of industry efficiency (12) will be:

$$IE^t = ITE^t + ISE^t + IME^t + IRE^t \tag{19}$$

The left-hand side is a measure of aggregate loss in outputs and waste in inputs at the industry level and is defined in equation (12). The right-hand side attributes this inefficiency to the different sources: ITE is inefficiency arising from individual firms technical inefficiencies (firms who do not operate on the production frontier); ISE is a measure of inefficiency deriving from individual firm size inefficiencies (i.e., firms operating on a too large scale); IME is a measure of potential efficiency gain from mergers of existing firms; IRE is accounting for possible reallocations of inputs and outputs among firms in order to move the industry configuration towards the efficient one. In decomposition (19) there is a substantial distinction between the first component and the last three components: while the first component arises naturally as an aggregation from the individual firm level, the last three components pertain to the way the industry is organized and are intrinsically related to the objective of the industry of providing an efficient way of delivering aggregate production. In fact, an increase in individual firm technical efficiency always leads to an increase in aggregate industry technical efficiency; on the contrary aggregate industry improvements in the last three components could come to the detriment of individual firm conditions (for example in a break-up overall profit of the firm may decrease). Therefore the last three indicators measure the extent to which the industry organization is efficient in transforming inputs into outputs. These indicators cannot be defined unless the industry perspective is brought into the analysis (i.e., these indicators can only be defined in aggregate terms). In order to give a more direct interpretation of the proportion to which the overall inefficiency of the industry depends on the different factors it is possible to normalize the right-hand side using the left hand side:

$$\%ITE^t + \%ISE^t + \%IME^t + \%IRE^t = 1 \tag{20}$$

where $\%ITE^t = \frac{ITE^t}{IE^t}$, $\%ISE^t = \frac{ISE^t}{IE^t}$, $\%IME^t = \frac{IME^t}{IE^t}$, $\%IRE^t = \frac{IRE^t}{IE^t}$.

A final remark has to be done on the decomposition of the industry organizational indicator embedded in (13). The aggregate definition of this indicator can be considered quite uncontroversial, since the two indicators which define it are very well indentified in equations (10) and (11). The decomposition of the industry organizational indicator (13) into the three components ISE, IME and IRE is more controversial and depends on the definition of break-ups and mergers here adopted. In fact, if one wants to be logically consequential, mergers and break-ups may be interpreted in terms of re-allocation of inputs and outputs among firms and therefore, from a production theory perspective, those two notions (and the associated indicators) are indistinguishable. For example, if firm A and B merge, the merger can be interpreted as a transfer of all the inputs and outputs of firm A to B (or viceversa), with A (or B) exiting the industry (which is also sometimes defined as an acquisition, but results in the same production unit of the merger). Similarly, a break-up of a single large firm into sub-units can be seen as the entry of those units into the industry and a re-allocation of inputs and outputs from the large firm to the new firms that entered the industry. Therefore all these break-ups and mergers may be thought as transfers of inputs and outputs among firms once entry and exit is allowed. We keep decomposition (19) as our benchmark because it is more directly related to the efforts the literature made to identify break-ups and mergers.

4. Dynamic: industry productivity change

All the previous discussion assumed a static point of view, focusing on a specific time period. Now, suppose there are two time periods, thus two data generated technologies T^t, T^{t+1} and two industry aggregate inputs and outputs $(\boldsymbol{I}^t, \boldsymbol{Q}^t)$, $(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1})$. Interestingly, the two time periods can also be interpreted as different groups of firms: one group is facing technology t and the other group technology t+1. With this group interpretation the way is also open to apply this proposed framework to meta-frontier analysis, where different groups of firms are facing different technologies. Industry performance can improve because of two main reasons: first, industry efficiency can increase between the two time periods and, second, the industry production frontier itself can shift in time (technical change). A measure of industry structural efficiency change can be obtained as the difference between the industry structural inefficiency indicators (11) in the two time periods:

$$IEC = IE^t - IE^{t+1} = D_A^t(\boldsymbol{I}^t, \boldsymbol{Q}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D_A^{t+1}(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1}; \boldsymbol{g}_x, \boldsymbol{g}_y) \qquad (21)$$

This indicator is higher (lower) than zero if industry efficiency increases (decreases). Since we established that the additive technology is the benchmark technology for the industry, technical change can be measured as the displacement of the enlarged additive production frontier at the point $(\boldsymbol{I}^t, \boldsymbol{Q}^t)$:

$$ITC^{t+1} = D_A^{t+1}(\boldsymbol{I}^t, \boldsymbol{Q}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D_A^t(\boldsymbol{I}^t, \boldsymbol{Q}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) \qquad (22)$$

This indicator is higher (lower) than zero if there is technical progress (regress). Using property P2 of the additive DDF this indicator can also be expressed as the displacement of the VRS technology at the point of the hypothetical sample average firm:

$$ITC^{t+1} = S_t^*[D^{t+1}(\bar{\boldsymbol{I}}^t, \bar{\boldsymbol{Q}}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D^t(\bar{\boldsymbol{I}}^t, \bar{\boldsymbol{Q}}^t; \boldsymbol{g}_x, \boldsymbol{g}_y)] \qquad (23)$$

In other words, this indicator measures how much the VRS production possibility set has expanded or contracted at the specific point $(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t)$ which has been identified as a benchmark from an aggregate industry perspective. The production possibility set can well collapse at other points (local technical regress) but the industry will still benefit from the technical change that happens at the point $(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t)$. Here it can be seen clearly that the additive technology is appropriate when benchmarking an industry: the VRS technology is the firm technology, while the industry has access to the additive technology (which is precluded to a single firm).

The sum of the efficiency change and the technical change components returns a measure of aggregate productivity change for the industry:

$$IProd^{t+1} = IEC + ITC^{t+1} = D_A^{t+1}(\boldsymbol{I}^t, \boldsymbol{Q}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D_A^{t+1}(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1}; \boldsymbol{g}_x, \boldsymbol{g}_y) \quad (24)$$

This indicator takes a value larger (smaller) than zero if industry productivity has improved (deteriorated). Once again the benchmark technology is the additive one because, from the industry point of view, the number of firms which operate in it is a variable. It should be noted that this indicator of productivity change corresponds to the comparison period Luenberger indicator introduced by Chambers et al(1996), evaluated at the point where the industry is operating and benchmarked against the additive technology (which is the industry technology). The ITC^{t+1} measure of technical change fixes the input-output industry vector at the base period value. Alternatively one could use the comparison period industry input-output quantities obtaining a different indicator of industry technical change:

$$ITC^t = D_A^{t+1}(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1}; \boldsymbol{g}_x, \boldsymbol{g}_y) - D_A^t(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1}; \boldsymbol{g}_x, \boldsymbol{g}_y) \quad (25)$$

and the associated base period Malmquist-Luenberger industry productivity indicator becomes:

$$IProd^t = IEC + ITC^t = D_A^t(\boldsymbol{I}^t, \boldsymbol{Q}^t; \boldsymbol{g}_x, \boldsymbol{g}_y) - D_A^t(\boldsymbol{I}^{t+1}, \boldsymbol{Q}^{t+1}; \boldsymbol{g}_x, \boldsymbol{g}_y) \quad (26)$$

As it is standard in this literature, to avoid the arbitrariness of choosing base or comparison period benchmarks, one can use the average of the previous Luenberger productivity indicators:

$$IProd = \frac{(IProd^t + IProd^{t+1})}{2} = ITC + IEC \quad (27)$$

Since a decomposition of industry efficiency has been provided in the previous section, it is possible to decompose industry inefficiency change (21) as:

$$IEC = ITEC + IOEC \quad (28)$$

where $ITEC = ITE^t - ITE^{t+1}$ and $IOEC = IOE^t - IOE^{t+1}$. Equation (28) avoids for simplicity the further decomposition of the industry organizational efficiency indicator IOE into size efficiency and merger efficiency. Inserting expression (28) into (27) provides an overall decomposition of industry productivity change:

$$IProd = ITC + ITEC + IOEC \quad (29)$$

The left-hand side of equation (29) is positive (negative) when the productivity of the industry increases (decreases). This increase (decrease) of productivity can be attributed to 3 different fundamental factors: *first*, local technological progress (regress) which implies an outward (inward) displacement of the industry technology; *second*, an overall increase (decrease) in the technical efficiency with which the firms in the industry are operating; *third*, a better (worse) configuration of the industry itself.

5. Shadow profit and the Lowe index of firm inefficiency

Chambers et al (1998) provided a duality theorem between the DDF and the profit function (see Fare and Primont, 2006 for a complete directional duality theory). Given a vector of output prices $\boldsymbol{p} \in \mathbb{R}_+^M$ and a vector of input prices $\boldsymbol{w} \in \mathbb{R}_+^N$ the profit function is defined as:

$$\Pi(\boldsymbol{p}, \boldsymbol{w}) = \sup_{x,y}\{\boldsymbol{py} - \boldsymbol{wx}: (\boldsymbol{x}, \boldsymbol{y}) \in T^t\} \tag{30}$$

and the following relationship is established between the profit function and the DDF:

$$\frac{\Pi(\boldsymbol{p},\boldsymbol{w})-(\boldsymbol{py}-\boldsymbol{wx})}{\boldsymbol{pg}_y-\boldsymbol{wg}_x} \geq D^t\left(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{g}_y, \boldsymbol{g}_y\right) \tag{31}$$

The left hand side is the Nerlovian profit efficiency indicator. If a vector of prices is available for the overall industry, then it makes sense to normalize the directional vector to be equal to one dollar value: $\boldsymbol{pg}_y - \boldsymbol{wg}_x = 1$. Since rescalings of the directional vector corresponds to rescalings of the DDF, this normalization is not substantive. With this normalization in place the inefficiency can be interpreted in physical terms (the number of units of each input and output expressed in a suitable unit of measurement) and can also be interpreted as a dollar loss measure.

If input and output prices are not available, then a shadow profit-maximizing problem will provide a vector of shadow prices. Shadow-prices are interpreted as prices which support a specific input-output bundle as profit maximizing; thus in the shadow profit maximization problem, prices are a variable and the inputs and outputs are fixed, giving rise to the dual of linear program (3) (see Fukuyama, 2003):

$$\Pi^*(\bar{\boldsymbol{I}}^t, \bar{\boldsymbol{Q}}^t) = \max_{\boldsymbol{p},\boldsymbol{w},\omega}\left\{\boldsymbol{p}\bar{\boldsymbol{Q}}^t - \boldsymbol{w}\bar{\boldsymbol{I}}^t - \omega : \boldsymbol{p}\bar{\boldsymbol{Q}}^t - \boldsymbol{w}\bar{\boldsymbol{I}}^t - \omega\boldsymbol{1}_K \leq \boldsymbol{0}_K, \boldsymbol{wg}_x + \boldsymbol{pg}_y = 1, \boldsymbol{p} \geq \boldsymbol{0}, \boldsymbol{w} \geq \boldsymbol{0}, \omega\ free\right\} \tag{32}$$

Let's call $(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{w}})$ the optimal solution to this problem. This vector of prices is interpreted as the hypothetical vector of prices which supports the observed efficient projection of the industry average total input and output as profit maximizing. The vector of shadow prices is a good benchmark also because, if implemented, it would lead profit maximizing firms to move to the efficient industry point, by means of mergers, break-ups and re-allocations, therefore maximizing the efficiency of production at the industry level. The shadow profit formulation (32) has been discussed by Fukuyama (2003) as the dual to the DDF primal formulation (3). Another way of interpreting this shadow prices is as follow: inserting the shadow prices into the original profit function (30) will return the optimal inputs and outputs associated with those shadow prices and this optimal bundle will be the industry average inputs and outputs $\left(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t\right)$.

These shadow prices can be used to build a shadow profit Nerlovian inefficiency measure at the firm level using a Lowe (fixed price) formula:

$$\hat{\pi}(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t) = \Pi^*\left(\bar{\boldsymbol{I}}_{Eff}^t, \bar{\boldsymbol{Q}}_{Eff}^t\right) - (\widehat{\boldsymbol{p}}\boldsymbol{y}_k^t - \widehat{\boldsymbol{w}}\boldsymbol{x}_k^t)$$

This is the total shadow profit inefficiency of firm k when compared to the optimal industry bundle $\left(\bar{\boldsymbol{I}}_{Eff}^{t}, \bar{\boldsymbol{Q}}_{Eff}^{t}\right)$. This overall shadow profit inefficiency is always larger than technical inefficiency thanks to the Mahler inequality:

$$\hat{\pi}(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t) \geq D^t\left(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y\right)$$

The difference between these two measures can be attributed to an activity effect (i.e. a mix of scale and scope economies). Therefore overall firm shadow profit efficiency can be decomposed in a technical inefficiency component and in an activity efficiency component:

$$\hat{\pi}(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t) = D^t\left(\boldsymbol{x}_k^t, \boldsymbol{y}_k^t; \boldsymbol{g}_x, \boldsymbol{g}_y\right) + AE$$

6. Data and empirical results

The data are drawn from the InfoBank and Casemix databases of Queensland Helath and comprises 116 public hospitals in the state of Queensland in the years 1996-2004 (balanced panel dataset). These data were elaborated by O'Donnell and Nguyen (2011) to whom the reader should refer for further details. There are three outputs (number of outpatient occasions, number of weighted episodes of surgical care, number of weighted episodes of medical care) and three inputs (full-time equivalent number of medical officers, full-time equivalent number of nurses, number of beds). Descriptive statistics are reported in table 1. The directional vector chosen for this empirical application is: $g_y = (0\ 0\ 0)$ and $g_x = (1\ 1\ 1)$. This means an input saving approach has been chosen, which corresponds to ask how much inputs (cost) can be saved keeping the level of output (treatments) constant.

Table 2 and 3 report the results of the industrial efficiency indicator along with its decomposition into the various components. Table 3 is particularly useful because it provides the decomposition in percentages terms. From this last table emerges quite clearly that technical efficiency represents a small component in the efficiency decomposition accounting for less than 15% of the total inefficiency of the industry across all the time periods. On the contrary the industry organizational indicator (IOE) (being always above 80% of total industrial inefficiency) accounts for the largest part of industrial inefficiency. The decomposition of the industrial organization indicator is quite interesting. Mergers and re-allocations of inputs and outputs account together for around half of the industry organizational inefficiency; the other half is accounted for by size inefficiencies. It should be emphasized once again that mergers and re-allocations are not always feasible (for example hospitals can be geographically distant); on the contrary size inefficiencies can be resolved by splitting a large hospital into a number of different administrative production units. Therefore the index of size inefficiencies (ISE) becomes particularly important in the light of the fact that a splitting is always physically possible (and the splitting does not necessarily mean that two different building must be constructed; it could mean that in the same building two different hospitals with different administrations are operating). Size inefficiency accounts for roughly 40% of total inefficiencies in the sector, pointing to the fact that a policy of administrative splitting of large hospitals could benefit the overall sector inefficiency, helping in reducing the cost of delivering health or increasing the number of treatments for a given cost.

Finally, table 4 reports the result of the productivity decomposition. Productivity growth has been slightly negative due to an implosion of the production frontier. O'Donnell and Nguyen (2011) found a similar result of negative technical change. This effect has been partially compensated by a slight increase in industry efficiency.

7. Conclusion

This study introduced a measure of industrial inefficiency. It has been shown that this measure can be defined only by taking an aggregate industry production perspective. Aggregation of firm's inefficiencies lead to an underestimation of the potential efficiency gains that an industry can realize thanks to organizational effects. Decomposing the industrial inefficiency indicator into different components the following effects have been identified: 1) inefficiencies arising from firms operating on a large size that can be split into smaller more productive units; 2) efficiency gains that can be realized thanks to the merger of small firms; 3) re-allocation of inputs and outputs in order to bring firms to the most optimal scale of production. The tool used to describe all these components was the directional distance function coupled with different types of data generated technologies. After defining the static industry inefficiency indicator, a dynamic decomposition of productivity change was proposed. Productivity change itself was decomposed into technical change and efficiency components. The methodology has been applied to data on hospitals showing that technical efficiency of individual hospitals accounts only for less than 15% of the total inefficiency of the industry. The most part of industry inefficiency has been found to be into the way the sector is organized, with a prominence of the size inefficiency component of around 40%. This last result in particular point to the fact that further investigation should be put into the causes that ingenerate a sub-additive technology for large hospitals and what is the incentive of hospitals to be on such an inefficient production scale.

References

Banal-Estanol A., Macho-Stadler I., Seldeslachts J. (2008), Endogenous mergers and endogenous efficiency gains: the efficiency defence revisited. International Journal of Industrial Organization, 26, 69-91

Banker R.D., Chang H., Natarajan R. (2005), Productivity change, technical progress and relative efficiency change in the public accounting industry, Management Science, 51, 291-304

Banker R.D., Charnes A., Cooper W.W. (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30, 1078-1092

Bogetoft P., Thorsen B.J., Strange N. (2003), Efficiency and merger gains in the Danish forestry extension service. Forest Science, 49 (4)

Bogetoft P., Wand D. (2005), Estimating the potential gains from mergers. Journal of Productivity Analysis, 23, 145-171

Briec W., Dervaux B., Leleu H. (2003), Aggregation of directional distance functions and industrial efficiency. Journal of Economics, 79, 237-261

Chambers R. G., Chung Y., Fare R. (1996), Benefit and distance functions. Journal of Economic Theory, 70, 407-419

Chambers R.G., Chung Y., Fare R. (1998), Profit, directional distance functions and Nerlovian efficiency. Journal of Optimization Theory and Applications, 98, 351-364

Fare R. (1986), Addition and efficiency. The Quarterly Journal of Economics, 101, 861-866

Fare R., Grosskopf S., Zelenyuk V. (2008), Aggregation of Nerlovian profit indicator. Applied Economics Letters, 15, 845-847

Fare R., Primont D. (2003), Luenberger productivity indicators: aggregation across firms. Journal of Productivity Analysis, 20, 425-435

Fare R., Primont D. (2006), Directional duality theory. Economic Theory, 29, 239-247

Fare R., Zelenyuk V. (2002), Input aggregation and technical efficiency. Applied Economics Letters, 9, 635-636

Fare R., Zelenyuk V. (2003), On aggregate Farrell efficiencies. European Journal of Operational Research, 146, 615-620

Given R.S. (1996), Economies of scale and scope as an explanation of merger and output diversification activities in the health maintenance organization industry. Journal of Health Economics, 15, 685-713

Green R. (1996), Increasing competition in the British electricity spot market. The Journal of Industrial Economics, 44, 205-216

Grifell-Tatje E., Lovell C.A.K. (1999), Profits and productivity, 45, 1177-1193

Kristensen T., Bogetoft P., Pedersen K.M. (2010), Potential gains from hospital mergers in Denmark. Health Care Management Science, 13, 334-345

Kwoka J., Pollitt M. (2010), Do mergers improve efficiency? Evidence from restructuring the US electric power sector. International Journal of Industrial Organization, 28, 645-656

Maindiratta A. (1990), Largest size-efficient scale and size efficiencies of decision-making units in data envelopment analysis. Journal of Econometrics, 46, 57-72

Nakamura T., Ohashi H. (2008), Effects of technology adoption on productivity and industry growth: a study of steel refining furnaces. The Journal of Industrial Economics, LVI, 470-499

Newbery D.M., Pollitt M.G. (1997), The restructuring and privatisation of Britain's CEGB – Was it worth it? The Journal of Industrial Economics, XLV, 269-303

O'Donnell C.J., Nguyen K. (2011), An econometric approach to estimating support prices and measures of productivity change in public hospitals. CEPA Working Paper, WP03/2011

Ray S.C. (2007), Are some Indian banks too large? An examination of size efficiency in Indian banking. Journal of Productivity Analysis, 27, 41-56

Ray S.C., Hu X. (1997), On the technically efficient organization of an industry: a study of U.S. airlines. Journal of Productivity Analysis, 8, 5-18

Ray S.C., Mukherjee K. (1998), Quantity, quality and efficiency for a partially super-additive cost function: Connecticut public schools revisited. Journal of Productivity Analysis, 10, 47-62

Sari N. (2003), Efficiency outcomes of market concentration and managed care. International Journal of Industrial Organization, 21, 1571-1589

Van Biesebroeck J. (2007), Robustness of productivity estimates. The Journal of Industrial Economics, LV, 529-569

Vasconcelos H. (2010), Efficiency gains and structural remedies in merger control. The Journal of Industrial Organization, 4, 742-766

Weber R.A., Camerer C.F. (2003), Cultural conflict and merger failure: an experimental approach. Management Science, 49, 400-415

Zelenyuk V. (2006), Aggregation of Malmquist productivity indexes. European Journal of Operational Research, 174, 1076-1086

Zhang F. (2007), Does electricity restructuring work? Evidence from the U.S. nuclear energy industry. The Journal of Industrial Economics, LV, 397-418

Zitzewitz E.W. (2003), Competition and long-run productivity growth in the UK and US tobacco industries, 1879-1939. The Journal of Industrial Economics, LI, 1-33

Tables and Figures

Table 1 – Descriptive Statistics

	Outpatients	Surgical	Medical	Medical officers	Nurses	Beds
Mean	64226.76	2276.90	3862.86	469.20	114.59	80.23
STD	121707.41	6340.46	7573.42	4119.70	257.52	148.22
Min	806.01	0.01	85.51	0.02	6.10	2.01
Max	1011976.00	44825.50	51536.27	80118.99	1864.19	1138.01

Table 2 – Industrial efficiency and its static decomposition

Year	IE (ITE+IOE)	ITE	IOE (ISE+IME+IRE)	ISE	IME	IRE
1996	1438.8	203.9	1234.9	698.3	177.8	358.8
1997	1762.7	258.4	1504.3	860.3	181.4	462.6
1998	1728.8	206.0	1522.8	782.9	242.2	497.7
1999	1605.9	237.0	1368.9	681.1	321.2	366.7
2000	1437.8	224.9	1212.9	598.5	247.7	366.8
2001	1313.9	183.1	1130.8	554.3	206.8	369.6
2002	1214.1	136.9	1077.2	491.0	117.3	468.8
2003	1228.0	103.8	1124.2	380.5	385.6	358.1
2004	2892.2	343.5	2548.6	646.2	1441.1	461.3

Table 3 – Percentage contribution to industrial efficiency by the different components

Year	IE (ITE+IOE)	ITE	IOE (ISE+IME+IRE)	ISE	IME	IRE
1996	100.0	14.2	85.8	48.5	12.4	24.9
1997	100.0	14.7	85.3	48.8	10.3	26.2
1998	100.0	11.9	88.1	45.3	14.0	28.8
1999	100.0	14.8	85.2	42.4	20.0	22.8
2000	100.0	15.6	84.4	41.6	17.2	25.5
2001	100.0	13.9	86.1	42.2	15.7	28.1
2002	100.0	11.3	88.7	40.4	9.7	38.6
2003	100.0	8.5	91.5	31.0	31.4	29.2
2004	100.0	11.9	88.1	22.3	49.8	16.0

Table 4 – Productivity growth and its decomposition

Year	Prod (ITC+IEC)	ITC	IEC (ITEC+IOEC)	ITEC	IOEC
1996/1997	-182.6	141.3	-323.9	-54.5	-269.4
1997/1998	-32.1	-66.0	33.9	52.4	-18.5
1998/1999	-119.6	-242.5	122.9	-31.0	153.9
1999/2000	-134.0	-302.0	168.0	12.0	156.0
2000/2001	-101.7	-225.6	124.0	41.9	82.1
2001/2002	19.8	-80.0	99.8	46.2	53.6
2002/2003	-31.0	-17.1	-13.9	33.2	-47.0
2003/2004	-2066.6	-402.4	-1664.2	-239.8	-1424.4