Benchmarking and Industry Performance

Thijs ten Raa*

June 2009

Abstract

Benchmarking is formalized by a linear program that determines the efficiency of a firm relative to its peers and is used to determine the efficiency of an industry. The overall efficiency is shown to be underestimated by mean firm efficiency and the bias is zero if and only if the firm shadow prices of the inputs and outputs generated by the benchmarking programs are equal across firms. Otherwise the bias provides an efficiency measure for the organization of the industry.

A main contribution of this paper is the interrelation of productivity analysis and the theory of industrial organization. A proposition proves that an industrial organization is efficient in the sense of productivity

*Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-4662365; fax: +31-13-4663280; e-mail: tenRaa@UvT.nl. I am grateful to Thanh Le Phuoc for good research assistance, William Weber for making his data set available, Shawna Grosskopf and Sverre Kittelsen for literature references, and Knox Lovell for telling me to analyze entry and exit.

analysis if and only if it is supportable in the entry-proofness sense of Sharkey and Telser (1978).

The known decomposition of performance in efficiency change and technical change is augmented with a term for the industrial organization efficiency change. The performance measure is shown to be consistent with the Solow residual and Malmquist indices for its components are given. An analysis of the Japanese banking industry illustrates and the dynamic effects of entry and exit can be accommodated.

JEL Classification Numbers: L10, D24, O47

Keywords: Scope (dis)economies, Industrial Organization, Efficiency, Aggregation, Productivity

1 Introduction

The theory of benchmarking can be traced back to Farrell (1957) and an easy introduction is given in Diewert's (2005) tutorial. The present, formal development connects benchmarking to the measurement of industrial organization (in)efficiencies and to economic performance measures. I define benchmarking through a function that maps two arguments, namely the object to be benchmarked–a firm–and the backdrop against which it is benchmarked–the industry–into a scalar, namely the efficiency of the firm. However simplistic, a firm can be represented by the pair of its input vector and its output vector. The mapping summarizes a program that determines how much more output the firm could produce with its inputs if it would reallocate them to the activities implicit in the other firms. The output maximizing reallocation identifies the best practice technologies or *benchmarks* for the firm under consideration and the inverse of the output expansion defines its *efficiency*.

The performance of a firm will be measured by its output/input ratio or *productivity*–a concept that I will connect explicitly to the aforementioned efficiency function. This performance measure may be increased by efficiency change or technical change. The first element reflects a relative movement of a firm in the direction of the best practice frontier while the latter element represents a shift of the frontier. Aggregate performance may rise more than firm performance scores suggest, if resources are better allocated between firms. The latter component is the industrial organization effect and this paper quantifies it.

Debreu (1951) recognized imperfections in economic organization as a source of aggregate inefficiency and Diewert (1983) analyzed them for open economies, focussing on price distortions. I operationalize these ideas, without using external price information (such as the world prices an open economy faces). In discrete time the operationalization will be expressed in terms of Malmquist indices, which are discrete time variants of productivity growth measures. Another novelty is that my theory of benchmarking extends my interrelationship of two value-based performance measures, namely Diewert's (1993) price index measurement and Solow's (1957) residual analysis, from the back-of-the-envelope calculation for production functions (ten Raa, 2005) to the activity framework.

If relatively productive firms grow relatively fast, the industry will improve its performance, even when no firm exhibits technical change or efficiency change. The insight is closely related to shift-share or (de)composition analysis (Denny, Fuss, Waverman ,1981 and Balk, 2003). For example, aggregating industry productivities to total productivity growth, Wolff's (1985, formula 23) decomposition analysis includes a composition effect, and Jorgenson, Ho, and Stiroh (2003, formula 53) capture allocative efficiency changes. Yet there is more to the performance of a composite such as an industry than micro or firm performance and composition effects. While this decomposition seems clear at the micro level and has been applied to the macro level (Färe et al., 1994), Blackorby and Russell (1999) and Briec, Dervaux and Leleu (2003) have shown that things do not add up except under restrictive conditions. Bartelsman and Doms' (2000, p. 571) understanding that "Aggregated productivity can be computed as the share-weighted average of individual productivity" is not exactly right. The first proposition of this paper will show that aggregated performance is *less*. There is a bias (ten Raa, 2005) and I will show that it reflects the industrial organization or, more precisely, the departure from an optimal organization. In a framework for the measurement of all performance components–technical change, efficiency change and industrial organization–I consolidate price and non-price based approaches, combining insights of composition analysis (which is neutral with respect to optimality issues) and of benchmarking (which is all about optimality but bears little on composition effects).

A main contribution of this paper is the interrelation of productivity analysis and the theory of industrial organization. A proposition proves that an industrial organization is efficient (in the sense of productivity analysis) if and only if it is supportable in Sharkey and Telser's (1978) sense of being entry-proof.

The efficiency/productivity literature boasts different econometric techniques and many econometricians feel strong about this. Roughly speaking the trade-off is between functional form limitations and robustness. In this theoretical paper, robustness is no issue however and it would be limiting and distracting to establish my propositions in a functional form framework. The exercise is one similar to Data Envelopment Analysis which is not robust–although there is recent bootstrapping work that effectively addresses the defect (Simar and Wilson, 2007)–but a functional form framework is less suitable for this theoretical paper.

The organization of the paper is as follows. The linear program is analyzed in section 2. Exploiting duality theory, I show that benchmarking is consistent with the price index approach to performance measurement. Much insight is gained by benchmarking not a firm but the entire industry. Section 3 uses the aggregation bias to measure the inefficiency of the industry. It relates this normative efficiency concept with the positive concept of supportability. Section 4 traces time changes of efficiency through both the object to be benchmarked and the reference benchmark, showing that the former transmits productivity growth and the latter technical change. Exploiting duality theory once more, I show that this dynamic benchmarking is consistent with the growth accounting concept of the Solow residual. Discrete time approximations are presented in section 5, including the industrial organization effect. Section 6 applies the theory to measure the evolution of the industrial organization of Japanese banking. Section 7 shows how the basic model can be modified to accommodate entry and exit of firms and other departures. Section 8 concludes.

2 Benchmarking: price analysis

Denote firm i's input vector by x^i and its output vector by y^i, $i = 1, ..., I$.[1] Input vectors have a common dimension, output vectors have a common dimension, and these two dimensions may differ. For example, inputs can be labor, capital, and land, while outputs may be numerous goods and services. Some commodities can be both input and output. However simplistic, an *industrial organization* can be represented by the allocation $(x^i, y^i)_{i=1,...,I}$, which is denoted briefly by (x, y). If $I = 1$, the industrial organization is a monopoly; if $I = 2$, it is a duopoly. If y is a diagonal matrix, we have monopolistic competition. If x is a row vector, we have an input price taking industry, for which inputs can be aggregated to 'cost.' The efficiency of a firm is determined by *benchmarking* its structure (x^i, y^i) against the industrial organization (x, y). This is a comparison between the actual output level and the best practice output level achievable with the available input vector, just as Farrell's (1957) productive efficiency measurement technique and Shephard's (1970) output distance function. The idea is to reallocate the input, x^i, over all the activities $j = 1, ..., I$, and to run the latter with intensities θ_j, as to inflate the output, y^i, by an expansion factor $1/\varepsilon$:

[1] All I need are firms' input and output data. This simplicity has a price: I miss performance failures on the demand side. However, a theoretical underpinning of this supply side approach is given in ten Raa (2008), where it is demonstrated that my supply side efficiency measure is conservative (in the sense of underestimating full inefficiency) and sharp (attained by certain demand functions).

$$\max_{\varepsilon,\theta_j\geq 0} 1/\varepsilon : \sum \theta_j x^j \leq x^i, y^i/\varepsilon \leq \sum \theta_j y^j \tag{1}$$

Here it is assumed that activity (x^i, y^i) can be run with constant returns to scale. An obvious extension of this paper will be the generalization to variable returns to scale, but this comes with integer problems (known in the theory of contestable markets) that obscure the relationship between the contestability and the efficiency of an industry revealed and measured in the present paper. In other words, that generalization better be relegated to a separate paper.

Instead of expanding output, one might contract the input, but under constant returns to scale output and input based benchmarking procedures are equivalent. So stick to (1). Let ε^i solve primal program (1).[2] The expanded y^i/ε^i is the *potential* output of firm i, using the best practice technologies. If $\varepsilon^i = 0.9$, firm i could produce a factor $1/0.9 = 1.11$ or 11% more. If $\varepsilon^i = 1$, potential output is no more than actual output and firm i is said to be fully efficient. In general, ε^i is a number between 0 and 1 which indicates the *firm efficiency* for firm i. The best practice firms or *benchmarks* relevant to firm i are signalled by $\theta_j > 0$ in program (1).

Denote the *shadow prices* of the constraints in (1) by w^i and p^i, for the inputs and outputs, respectively; they solve the dual program:[3]

$$\min_{p,w\geq 0} wx^i : py^j \leq wx^j, py^i = 1, \text{ all } j \tag{2}$$

[2] Notice the program is linear in the nonnegative variables $1/\varepsilon, \theta_j$.

[3] The price normalization constraint features no slack, because the non-negativity constraint for $1/\varepsilon$ is non-binding (as $1/\varepsilon = 1$ is feasible by choice of $\theta_i = 1$ and $\theta_j = 0, j \neq i$).

This dual program is essentially the original tack to Data Envelopment Analysis, taken by Charnes, Cooper and Rhodes (1979). Under the alternative scheme of input contraction, the dual would maximize the value of output. Either way the dual variables weigh output and input scores, which is particularly useful in situations where weights are not readily given, as in the assessment of EU internal market dynamics by Cherchyea, Lovell, Moesen, and Van Puyenbroeck (2007). The connection between efficiency and valuations–a central theme in this paper–is made by the main theorem of linear programming. According to this theorem the primal and dual programs have equal solution values:

$$1/\varepsilon^i = w^i x^i \tag{3}$$

Substituting the price normalization constraint of program (2) in equation (3), the efficiency of firm i becomes:

$$\varepsilon^i = p^i y^i / w^i x^i \tag{4}$$

This result establishes a connection between benchmarking and price index measurement. The efficiency is the *ratio* of the value of output to the value of input at shadow prices. Georgescu-Roegen (1951) called it return to the dollar in a more complicated framework, involving investment. Under constant reurns to scale the ratio should equal unity for profit maximizers. The ratio should be contrasted with the difference concept of Glass, McCallion, McKillop, and Stringer (2006), who argue that 'profit' as the difference between revenue and cost, provides a measure of publicly funded institutions' financial performance. Another perhaps surprising feature is that the ratio is evaluated at *private* prices, because the potential output of firm i has idiosyncratic commodity proportions and because of the presence of multiple inputs. If the output mix of a firm is relatively intensive in terms of some input, the shadow price of that input will be high. If there were essentially one input, as for an industry that is input price taking, then the shadow prices of the outputs can be shown to be independent of the firm's output mix (the Samuelson substitution theorem, ten Raa, 1995), ensuring perfect agreement between private and social values.

Shadow input prices are high.[4] Indeed, by the dual constraints in program (2), no firm makes positive profit at shadow prices. Benchmarks break even (by the phenomenon of complementary slackness) and inefficient firms incur a loss. This observation confirms that efficiency measure (4) is a number between 0 and 1.

We now apply the apparatus to the efficiency of the industry. Farrell (1957) and Førsund and Hjalmarsson (1979) call it structural efficiency and Johansen (1972) defines potential industry output as a function of total input. Following Färe and Grosskopf's (2004) extension, the idea is to reallocate the combined inputs of all firms, industry input $\overline{x} = \sum x^i$, as to inflate the aggregate output vector, $\overline{y} = \sum y_i$, by an expansion factor $1/\varepsilon$. Under constant returns to scale the industry technology is the cone spanned by the firm technologies, which is

[4] This is paradoxal, since the dual program minimizes the value of the inputs. The constraint, however, lifts the value of the inputs over the value of the outputs, in fact to the level of potential output. And the latter is maximized by the primal program.

the same technology as the reference technology in the firm efficiency program (1). The only modification to the latter is, therefore, the replacement of firm resources and output target by industry resources $\overline{x}$ and output target $\overline{y}$:

$$\max_{\varepsilon,\theta_j\geq 0} 1/\varepsilon : \sum \theta_j x^j \leq \overline{x}, \overline{y}/\varepsilon \leq \sum \theta_j y^j \tag{5}$$

Let $\overline{\varepsilon}$ solve program (5). It is a number between 0 and 1 which indicates the *industry efficiency*. The best practice firms or *benchmarks* relevant to the industry are signalled by $\theta_j > 0$ in program (5). Denote the shadow prices of the constraints in program (5) by $\overline{w}$ and $\overline{p}$, for the inputs and outputs, respectively; these *industry prices* solve the dual program:

$$\min_{p,w\geq 0} w\overline{x} : py^j \leq wx^j, p\overline{y} = 1, \text{ all } j \tag{6}$$

Analogous to equation (3), potential output increases by the following factor:

$$1/\overline{\varepsilon} = \overline{wx} \tag{7}$$

Analogous to equation (4), industry efficiency becomes:[5]

$$\overline{\varepsilon} = \overline{py}/\overline{wx} \tag{8}$$

Again, the efficiency is the ratio of the value of output to the value of input.

[5]Since prices are in the numerator and the denominator of formula (8), the price normalization in (6) is a wash.

3 Industrial organization: an efficiency measure

Proposition 1 establishes a one-sided relationship between the efficiencies of the firms and the efficiency of the industry. The gap in the relationship will be used to quantify the efficiency of the industrial organization.

Proposition 1. Industry efficiency is *less* than the market share weighted harmonic mean of the firm efficiencies: $\overline{\varepsilon} \leq 1/\sum \frac{s^i}{\varepsilon^i}$, where $s^i = \overline{p}y^i/\overline{py}$ are the market shares evaluated at the prices determined by dual program (6).

Proof. In the dual program (2), consider the socially optimal prices $\overline{p}/\overline{p}y^i$ and $\overline{w}/\overline{p}y^i$ (which need *not* be privately optimal). The denominator has been chosen as to fulfil the price normalization constraint in program (2) and the inequality constraint carries over from program (6). In short, these prices are feasible with respect to program (2). But by their suboptimality (in this private minimization program), $(\overline{w}/\overline{p}y^i)x^i \geq w^i x^i$ or $\overline{w}x^i \geq \overline{p}y^i w^i x^i = \overline{p}y^i/\varepsilon^i$, using equation (3). Summing and invoking equation (7) and the price normalization constraint of (6), we obtain $1/\overline{\varepsilon} = \overline{wx} = \overline{w}\sum x^i \geq \sum \overline{p}y^i/\varepsilon^i = \sum s^i/\varepsilon^i$. Inverting, industry efficiency becomes $\overline{\varepsilon} \leq 1/\sum \frac{s^i}{\varepsilon^i}$. Q.E.D.

The next proposition shows that the gap is zero if and only if the private prices of the firms coincide. The sufficiency part is reminiscent of Koopmans (1957) result that the industry profit function equals the sum of the firm profit functions.[6] Proposition 2 will be used later to show that an industrial organization is efficient if and only if it features price competition and free entry.

Proposition 2. The industry efficiency *equals* the market share weighted harmonic mean of the firm efficiencies if and only if all relative private price vectors match.

Proof. The private prices p^i and w^i solve program (2). Rescale them by letting them solve $\min_{p,w\geq 0} wx^i : py^j \leq wx^j, p\overline{y} = 1$, all j (normalizing against $\overline{y}$ instead of y^i). If the relative price vectors are equal, i.e. (p^i, w^i) are collinear, then, by the now common normalization constraint, (p^i, w^i) are equal, say $(\overline{p}, \overline{w})$. I claim these prices solve program (6). Feasibility carries over from any of the rescaled linear programs. Suppose the prices are not optimal. Then there would be a superior (p^*, w^*) with $p^*y^j \leq w^*x^j, p^*\overline{y} = 1$ and $w^*\overline{x} < \overline{wx}$. Hence $w^*x^i < \overline{w}x^i$ for some i. This contradicts that $(\overline{p}, \overline{w})$ solves the rescaled dual program of that firm i. Since $(\overline{p}, \overline{w})$ are also privately optimal, the inequalities in the Proof of Proposition 1 are binding. The bindingness of the last of these inequalities equates industry efficiency with the harmonic mean of the firm efficiencies. Conversely, if this equality holds, the inequalities in the Proof of Proposition 1 are binding. Hence $\overline{p}/\overline{p}y^i$ and $\overline{w}/\overline{p}y^i$ are privately optimal. These private prices are collinear. Q.E.D.

[6] Summation of profit functions presumes price equality across firms.

The reason that industry efficiency is less than mean firm efficiency is that the industrial organization is suboptimal. It is a form of allocative inefficiency. Firms better be split or merged, specialize or diversify. The optimal industrial organization is determined by the benchmarks in program (5). Suboptimality is signalled by a distortion between private and social prices (Proposition 2). The efficiency of the industrial organization can thus be measured by the ratio of the industry efficiency to the mean firm efficiency, or, using Proposition 1:

Definition 1. The *efficiency of an industrial organization*, (x, y), equals $\varepsilon^{IO} = \overline{\varepsilon} \sum s^i/\varepsilon^i$, where s^i are the market shares evaluated at the prices determined by dual program (6).

Notice that by Proposition 1 the efficiency of an industrial organization is a number between 0 and 1, with the latter value representing full efficiency according to Proposition 2.

Examples. 1. Consider an industry with equally efficient firms: $\varepsilon^i = \varepsilon$. Then by Proposition 1, $\overline{\varepsilon} \leq 1/\sum \frac{s^i}{\varepsilon^i} = \varepsilon$. Hence industry efficiency is less than firm efficiency. The efficiency of the industrial organization is $\varepsilon^{IO} = \overline{\varepsilon}/\varepsilon$.

2. Consider an industry that produces a single good from labor and capital. Three firms each produce one unit of output. Firm 1 uses just one unit of labor, firm 2 uses just one unit of capital, and firm 3 uses 1/3 units of both inputs. Since firm 1 has labor only, the technologies of firms 2 and 3 (which employ capital) are of no use. There is no potential increase of its output. The same conclusion holds for firm 2. Firm 3 could reallocate its labor and capital to the technologies employed by firms 1 and 2, respectively, but its output would go down from 1 to 2/3. Hence no firm has scope for an increase in output. All potential outputs are equal to the observed outputs and, therefore, all firms are 100% efficient. The industry, however, is *not* efficient. If firms 1 and 2 would merge and adopt the technology of firm 3, the new firm would be three times as big as firm 3, hence produce three units of output, which is one more than they produce using their own technologies. Potential output is four units (instead of three), so that the expansion factor is 4/3 and, therefore, the industry efficiency is 3/4 or only 75%. The efficiency of the industrial organization is 75/100 = 0.75 or 75%. The industry would do better if the two specialized firms would merge.

3. It is straightforward to construct an example where the industry would do better if a firm were split: Simply substitute diseconomies of scope for the economies of scope in Example 2, by letting firm 3 use 2/3 units of both inputs (instead of the 1/3 in Example 2).

4. Add a fourth firm to Example 2, with the same inputs as firm 3, but only 1/2 a unit of output. Clearly, firm 4 could produce a full unit of output (adopting the technology of firm 3). Its efficiency is 50%. In the present example, the outputs are 1, 1, 1, 0.5. The market shares are 2/7, 2/7, 2/7, 1/7. The firm efficiencies are 100%, 100%, 100%, 50%. The harmonic mean is $1/\sum s^i/\varepsilon^i = 1/(\frac{2/7}{1.00} + \frac{2/7}{1.00} + \frac{2/7}{1.00} + \frac{1/7}{0.50})$ or 87.5%. For the industry potential output is three for firms 1 and 2 jointly (see Example 2) and one for firms 3 and 4 each, hence five in total (instead of three and a half), so that the expansion factor is 5/3.5 and, therefore, the industry efficiency is 3.5/5 or only 70%. The efficiency of the industrial organization is $70/87.5 = 0.8$ or 80%.

An immediate consequence of Definition 1 is the following.

Proposition 3. Industry efficiency is the product of (market share weighted harmonic) mean firm efficiency and the efficiency of the industrial organization.

Proposition 3 will enable us to refine the decomposition of productivity growth in technical change and efficiency change in the next section, but first I make an observation on the connection between the efficiency and the supportability of an industrial organization in a contestable market.

In a contestable market, a potential entrant can tap the incumbents' technology and the entry costs are zero; see Baumol, Panzar and Willig (1982). Their solution concept is that of a sustainable industrial organization. The definition of sustainability involves demand conditions, but a simple, necessary condition is Sharkey and Telser's (1978) supportability, and that is all I need.[7]

[7] Students of the history of economic thought recognize this as another instance where the returns to scale (which are implicitly constant in benchmarking) determine if values are determined by the supply side or in interaction with the demand side: the classical and neo-classical points of view.

Definition 2. An industrial organization $(x^i, y^i)_{i=1,\ldots,I}$ is *supportable* by price vector (p, w) if the latter renders every incumbent firm profitable ($wx^i \leq py^i$), but any potential entrant unprofitable ($\sum \theta_j x^j \leq x^e, y^e \leq \sum \theta_j y^j, \theta_j \geq 0 \Rightarrow wx^e \geq py^e$).

The next result connects the benchmark industrial organization involving price competition plus free entry with efficient outcomes. In fact, Proposition 4 articulates the Baumol and Fischer (1978, p. 461) intuition that "it would be highly surprising if there were not a rough correspondence between the most economical market form and what actually occurs" and extends Baumol, Panzar and Willig's (1982) proposition that in a sustainable industry configuration total industry cost must be minimized.

Proposition 4. An industrial organization is supportable if and only if it is fully efficient and comprises fully efficient firms.

Proof. Let me first prove the only if part, which is known, at least for given input prices. So let the industrial organization be supportable by (p, w). I claim that the normalized supporting prices p/py^i and w/py^isolve the dual program of an incumbent firm, (2). For this purpose, take $\theta_j = 1$ and $\theta_k = 0, k \neq j$ in Definition 2. Then $x^j \leq x^e, y^e \leq y^j \Rightarrow wx^e \geq py^e$. In particular, $wx^j \geq py^j$. (In fact, the inequality is binding by the first condition in Definition 2.) This shows that the first constraint in (2) is met. The second constraint is met by the normalization by py^i. The value of the objective function in (2) must be at least $wx^i \geq py^i$ according to the first constraint in (2) with $j = i$. By the first condition in Definition 2 this lower bound py^i is not exceeded by wx^i of Definition 2, which renders the latter optimal in (2).

By equation (4), firm i is fully efficient. Moreover, since the relative prices solving dual program (2) are common to all incumbent firms, Proposition 2 applies and the industry efficiency is also 1. Substituting these findings in Definition 1, it follows that the industrial organization must be fully efficient.

To prove the converse, consider a fully efficient industrial organization comprising fully efficient firms. By Proposition 2, there is a common price vector $(\overline{p}, \overline{w})$ such that $\overline{p}/\overline{p}y^i$ and $\overline{w}/\overline{p}y^i$ are privately optimal. I claim that $(\overline{p}, \overline{w})$ supports the industrial organization. By equation (4), $\overline{w}x^i = \overline{p}y^i$. Let $\sum \theta_j x^j \leq x^e$ and $y^e \leq \sum \theta_j y^j, \theta_j \geq 0$. Then $wx^e \geq w \sum \theta_j x^j \geq p \sum \theta_j y^j \geq py^e$, where the middle inequality holds term by term in view of the constraints in program (2) that characterizes the privately optimal prices. Q.E.D.

4 Productivity growth: the recovery of the Solow residual

Extending ten Raa's (2005) back-of-the-envelope calculation for production functions, I will now connect the present activity framework with neoclassical growth accounting, particularly the Solow (1957) residual. This residual is the difference between the output and input growth rates and measures technical change in a competitive environment. Both growth rates are share-weighted expressions, using competitive valuations. In a noncompetitive environment the residual also captures efficiency change. Since the latter concept is defined by the benchmarking program (1), which features no prices, a little work has to be done to make the connection. Introduce time by subscripting inputs and outputs, as well as the derived constructs, using the symbol t. Firm i has input and output vectors x_t^i and y_t^i, $i = 1, ..., I$. By benchmarking, y_t^i/ε_t^i is derived, the potential output of firm i. Its efficiency is indicated by ε_t^i, a number between 0 and 1. As a percentage, efficiency change is:

$$EC_t^i = \frac{d}{dt}\varepsilon_t^i/\varepsilon_t^i \tag{9}$$

To bring in the technical change component of productivity growth, consider two polar cases. First, if the production possibilities of the industry remain constant, but firm i improves its output/input ratio (or productivity), approaching the production possibility frontier, the better internal organization yields positive efficiency change, while technical change is zero. Second, if firm i has a constant output/input ratio, but the production possibility frontier of the industry shifts out, the better external organization implies negative efficiency change for the firm. In this case efficiency change and technical change cancel and the firm has zero productivity growth. In either case, efficiency change and technical change sum to the firm's productivity growth:

$$PG_t^i = EC_t^i + TC_t^i \tag{10}$$

At this junction equation (10) is intuitive. Efficiency change has been defined, in equation (9), but technical change and productivity growth not yet and the equation must be demonstrated. Technical change manifests itself as a shift of the production possibility frontier. At each point of time, the frontier is determined by the industrial organization (x_t, y_t). The efficiency of firm i is determined by program (1). Its input-output pair, (x_t^i, y_t^i), is benchmarked against (x_t, y_t). Formally, program (1) determines the efficiency of firm i as a function of (x_t^i, y_t^i) and (x_t, y_t). Hence we may write $\varepsilon_t^i = e((x_t^i, y_t^i), (x_t, y_t))$, where mapping e summarizes the efficiency program. Now the program that determines the efficiency of the industry, (5), has *precisely* the same structure as that for the firms (the only difference is that the pair of *total* input and output vectors replace the role of a firm's pair), hence the *same* mapping e governs the relationship between the data and industry efficiency. The only difference is that it benchmarks the industry input-output pair, $(\overline{x}_t, \overline{y}_t)$. Consequently, program (5) may be written as $\overline{\varepsilon}_t = e((\overline{x}_t, \overline{y}_t), (x_t, y_t))$, with the same mapping e. Mapping e has two arguments, the input-output pair that is benchmarked, (x_t^i, y_t^i) in case of the firm, and the industry constellation that determines the frontier, (x_t, y_t). The structure of program (1) or (5) is independent of time, so that time does not enter the function as a separate argument. Denote the two partial derivatives of the mapping by e_1 and e_2.[8] By total differentiation, the efficiency change of firm i is:

$$\begin{aligned} &\tfrac{de}{dt}((x_t^i, y_t^i), (x_t, y_t))/e((x_t^i, y_t^i), (x_t, y_t)) \\ &= e_1((x_t^i, y_t^i), (x_t, y_t))\tfrac{d}{dt}(x_t^i, y_t^i)/e((x_t^i, y_t^i), (x_t, y_t)) \\ &+ e_2((x_t^i, y_t^i), (x_t, y_t))\tfrac{d}{dt}(x_t, y_t)/e((x_t^i, y_t^i), (x_t, y_t)) \end{aligned} \tag{11}$$

The measurement of technical change is subtle. If firm i stays put–$(x_t^i, y_t^i) =$ constant–but potential output increases, there must be technical progress. Now an increase in potential output, $1/\varepsilon_t^i$, is equivalent to a decrease in efficiency, ε_t^i. Hence a *negative* second partial derivative–which captures the *external* effect–indicates technical progress. Indeed, if a firm is fixed, but it becomes less efficient, it must be that the benchmarks have moved out. In short, *technical change* is measured by:

$$TC_t^i = -e_2((x_t^i, y_t^i), (x_t, y_t))\frac{d}{dt}(x_t, y_t)/e((x_t^i, y_t^i), (x_t, y_t)) \tag{12}$$

Finally, *productivity growth* of firm i ought to be defined irrespective the shift of the production possibility frontier. Productivity growth is defined by the effect of its *own* inputs and outputs on the efficiency performance of the firm:

$$PG_t^i = e_1((x_t^i, y_t^i), (x_t, y_t))\frac{d}{dt}(x_t^i, y_t^i)/e((x_t^i, y_t^i), (x_t, y_t)) \tag{13}$$

[8] Both are row vectors, because either argument has a number of components: the number of commodities and the I-fold of the number of commodities, respectively. The partial derivatives capture the shadow prices, as revealed by the ensuing analysis.

The internal versus external approaches to the concepts of productivity growth and technical change, respectively, is consistent with Solow's (1957) *residual* of firm i, which is defined by $p_t^i \frac{d}{dt} y_t^i / p_t^i y_t^i - w_t^i \frac{d}{dt} x_t^i / w_t^i x_t^i$. The reason is that our partial derivatives of the efficiency function involves the pricing of input and output changes according to the marginal products of the firm. More precisely, Proposition 5 proves that expression (13) equals the Solow residual. It completes the connect between the benchmarking and neoclassical approaches to productivity.

Proposition 5. Productivity growth is measured by the Solow residual: $PG_t^i = p_t^i \frac{d}{dt} y_t^i / p_t^i y_t^i - w_t^i \frac{d}{dt} x_t^i / w_t^i x_t^i$.

Proof. The proof is by duality analysis. Mapping e's first (sub)argument, x_t^i, lists the bound in program (1). The partial derivative of the objective value, $1/\varepsilon_t^i$, with respect to this bound is the shadow price w_t^i. By the chain rule, $\frac{-1}{\varepsilon_t^{i2}} \frac{\partial \varepsilon_t^i}{\partial x_t^i} = w_t^i$. Mapping e's next (sub)argument, y_t^i, is a coefficient in program (1). Setting up the Lagrangian function and application of the envelop theorem yields that the partial derivative of the objective value, $1/\varepsilon_t^i$, with respect to coefficient y_t^i is $-p_t^i(1/\varepsilon_t^i)$. By the chain rule, $\frac{-1}{\varepsilon_t^{i2}} \frac{\partial \varepsilon_t^i}{\partial y_t^i} = -p_t^i(1/\varepsilon_t^i)$. Substituting these in expression (13), we obtain $PG_t^i = (-w_t^i \varepsilon_t^{i2}, p_t^i \varepsilon_t^i) \frac{d}{dt}(x_t^i, y_t^i)/\varepsilon_t^i = (p_t^i \frac{d}{dt} y_t^i - \varepsilon_t^i w_t^i \frac{d}{dt} x_t^i)/p_t^i y_t^i = p_t^i \frac{d}{dt} y_t^i / p_t^i y_t^i - w_t^i \frac{d}{dt} x_t^i / w_t^i x_t^i$, by the price normalization constraint of program (2) and equation (4). Q.E.D.

Summarizing, efficiency change is defined by (9), technical change by (12), productivity growth by (13), and the former two sum to the latter by equation (11), which confirms our intuitive equation (10).

Things look only slightly different at the level of the industry. Now industry input and output, $(\overline{x}_t, \overline{y}_t)$, are benchmarked against the frontier. The productivity growth of the industry is:

$$\overline{PG}_t = e_1((\overline{x}_t, \overline{y}_t), (x_t, y_t)) \frac{d}{dt}(\overline{x}_t, \overline{y}_t) / e((\overline{x}_t, \overline{y}_t), (x_t, y_t)) \tag{14}$$

This expression is basically a summation of the firm productivity growth rates, (13), with the modification that private shadow prices have been replaced by social values. This difference constitutes precisely the aggregation bias uncovered by ten Raa (2005). The same difference between private and social valuations causes a bias in the aggregation of technical change, (12), but here it is a minor phenomenon, specific to the nonparametric approach. Consequently, the productivity aggregation bias is basically equal to the efficiency aggregation bias, or, invoking Proposition 3, the industrial organization effect. The next section will explain the role of industrial organization in the performance measure of productivity.

5 Malmquist indices

In discrete time a fascinating thought construct is to benchmark a firm against the industry at a different period. One may hope that the efficiency of a firm benchmarked against the industry in the next period, $e((x_t^i, y_t^i), (x_{t+1}, y_{t+1}))$, is low. The basic idea of the Malmquist productivity index is to trace firm i from period t to $t+1$ and to measure the change in efficiency relative to a fixed benchmark. For example, benchmarking against the second period yields

$e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1})) - e((x_t^i, y_t^i), (x_{t+1}, y_{t+1}))$. This difference expression is a discrete time version of the numerator of productivity growth expression (13). Efficiency change contributes to the first term and technical change to the second term. The discrete time frame prompts two minor modifications. First, Malmquist indices are ratios instead of level changes, so that the difference expression turns $\frac{e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1}))}{e((x_t^i, y_t^i), (x_{t+1}, y_{t+1}))}$.[9] Second, Caves, Christensen, and Diewert (1982) point out that one could just as well benchmark against the previous period, which would yield $\frac{e((x_{t+1}^i, y_{t+1}^i), (x_t, y_t))}{e((x_t^i, y_t^i), (x_t, y_t))}$. Färe et al. (1989) resolve this multiplicity by taking the geometric average of the two possibilities is taken. In short, the *Malmquist productivity index* is defined by:

$$M_t^i = \sqrt{\frac{e((x_{t+1}^i, y_{t+1}^i), (x_t, y_t))}{e((x_t^i, y_t^i), (x_t, y_t))} \cdot \frac{e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1}))}{e((x_t^i, y_t^i), (x_{t+1}, y_{t+1}))}} \tag{15}$$

Färe et al. (1989) decompose the index (15) in efficiency change and technical

[9] At a first order Taylor approximation this is equivalent to a change of variable, to the exponential function. For example, a growth rate of 1% will yield an index of 1.01. It is customary though to stick to the percentage format in reporting Malmquist indices.

change as follows:

$$M_t^i = \frac{e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1}))}{e((x_t^i, y_t^i), (x_t, y_t))} \cdot \sqrt{\frac{e((x_t^i, y_t^i), (x_t, y_t))}{e((x_t^i, y_t^i), (x_{t+1}, y_{t+1}))} \cdot \frac{e((x_{t+1}^i, y_{t+1}^i), (x_t, y_t))}{e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1}))}} \tag{16}$$

The first quotient in decomposition (16) measures the increase in efficiency from time t to time $t+1$. The remainder (the square root) contains two quotients in which the firm is fixed (at time t, respectively $t+1$), but the benchmark moves; this measures technical change.

Turning from firm i to the industry, benchmark industry input and output against the frontier. Comparison with the firm index (15) shows that the inudstry Malmquist productivity index becomes:

$$\overline{M}_t = \sqrt{\frac{e((\overline{x}_{t+1}, \overline{y}_{t+1}), (x_t, y_t))}{e((\overline{x}_t, \overline{y}_t), (x_t, y_t))} \cdot \frac{e((\overline{x}_{t+1}, \overline{y}_{t+1}), (x_{t+1}, y_{t+1}))}{e((\overline{x}_t, \overline{y}_t), (x_{t+1}, y_{t+1}))}} \tag{17}$$

Proposition 6. The Malmquist productivity index aggregates the change in the efficiency of the industrial organization, firm efficiency changes, and technical change:

$$\overline{M}_t = \frac{\varepsilon_{t+1}^{IO}}{\varepsilon_t^{IO}} \cdot \frac{\sum s_t^i / e((x_t^i, y_t^i), (x_t, y_t))}{\sum s_{t+1}^i / e((x_{t+1}^i, y_{t+1}^i), (x_{t+1}, y_{t+1}))} \cdot$$
$$\cdot \sqrt{\frac{e((\overline{x}_t, \overline{y}_t), (x_t, y_t))}{e((\overline{x}_t, \overline{y}_t), (x_{t+1}, y_{t+1}))} \cdot \frac{e((\overline{x}_{t+1}, \overline{y}_{t+1}), (x_t, y_t))}{e((\overline{x}_{t+1}, \overline{y}_{t+1}), (x_{t+1}, y_{t+1}))}}$$

The first quotient measures the change in the efficiency of the industrial organization. Firm efficiencies are aggregated in the second quotient by the market share weighted harmonic mean and market shares are evaluated at the shadow prices of the industry efficiency program (5). The square root measures technical change.[10]

Proof. Apply formula (16) to the industry and substitute, using Definition 1, for $e((\overline{x}_t, \overline{y}_t), (x_t, y_t)) = \varepsilon_t^{IO} / \sum s_t^i / e((x_t^i, y_t^i), (x_t, y_t))$ and similar for $e((\overline{x}_{t+1}, \overline{y}_{t+1}), (x_t, y_t))$. Q.E.D.

6 Illustration

Consider the Japanese banks ($i = 1, ..., I = 136$) over a five year period ($t = 1992, ..., 1996$).[11] There are three inputs (labor, capital, and funds from customers) and two outputs (loans and other investments). Formally we have a panel of inputs and outputs, (x_t^i, y_t^i). For the four transitions between periods Thanh Le Phuoc has computed the dynamic performance measure of productivity growth, and, applying Proposition 3, its decomposition in the industrial organization effect, firms efficiency change and technical change. The results are in Table 1.[12]

[10] In view of the previous note, we essentially have the *sum* of the three effects.

[11] Fukuyama and Weber (2002) kindly made available their data. The data were obtained by extracting Nikkei's data tape of bank financial statements. Six banks had missing data and were excluded. These were Akita Akebono, Bank of Tokyo, Hanwa, Hyogo, Midori, and Taiheiyo.

[12] When Malmquist indices and its components are properly reported as fractions of the order 1, the product of the components equals total productivity growth. When reported as percentages, the components sum to total factor productivity growth, up to a first order

Table 1.

The three contributions to the performance of the Japanese banking industry

Period	Industrial Organization	Firms Efficiency	Technical Change	Total Productivity
1992-1993	0.07%	-0.51%	0.52%	0.08%
1993-1994	0.57%	0.31%	-0.58%	0.29%
1994-1995	-0.45%	0.35%	1.28%	1.18%
1995-1996	0.23%	0.70%	2.19%	3.15%
Total, annualized	0.11%	0.21%	0.85%	1.17%

The results permit a diagnosis of the Japanese banking industry. In the mid 1990s Japanese banking showed a solid performance of 1.17% productivity growth per year, much due to a final sprint. The bulk was due to technical change, think of advances in electronic banking. The second biggest chunk was due to efficiency change at the bank level, think of the spread of ATMs. Last and least, the industrial reorganization contributed to the Japanese banking productivity growth. Various explanations can be advanced to understand these different contributions, such as R&D, competitive pressure, and changes

Taylor approximation. This and rounding errors explain why not all row figures add.

in bankruptcy procedures. Many observers feel that there is scope for a bigger role of the industrial reorganization of Japanese banking. True or not, the first task seems to be the measurement of its share in productivity growth and that can now be ticked off the research agenda.

7 Entry and exit and other departures of the model

An important source of productivity growth is the entry and exit of relatively productive and relatively unproductive firms (Bartelsman and Doms, 2000). Diewert and Fox (2005) discuss and illustrate the decomposition problems. My approach resolves some of their issues, since prices are endogenous. However, the aggregation bias issue must be handled.

An analysis of entry and exit requires the consideration of different numbers of firms at the beginning and end of a period. Thus, denote the number of firms at time t by I_t. Moreover, we must consider entrants and exitors. There is a slight time asymmetry. The novelty of entrants' technologies means they were not available in the past, but the oldness of exitors' technologies does not mean they are no longer available in the future. Old technologies become obsolete economically (unprofitable). Tulkens and Vanden Eeckaut (1995) model this asymmetry by means of sequential Malmquist indices. While entrants are modelled as new firms, exitors continue to exist, but become dormant–with zero output and input levels.

Denote the number of entrants by E_t. Then $I_{t+1} = I_t + E_t$ and the partition of firms at time $t+1$ in incumbents and entrants reads $\mathcal{I}_{t+1} = \mathcal{I}_t \cup \mathcal{E}_t = \{1, ..., I_t, I_t + 1, ..., I_t + E_t\}$. Denote the total input-output combinations of incumbents by $(\overline{x}_{t+1}, \overline{y}_{t+1})^{\mathcal{I}} = (\overline{x}^{\mathcal{I}}_{t+1}, \overline{y}^{\mathcal{I}}_{t+1})$ and similar for the entrants. Similarly, the industrial organization at time $t+1$ can be written $(x_{t+1}, y_{t+1}) = ((x_{t+1}, y_{t+1})^{\mathcal{I}}, (x_{t+1}, y_{t+1})^{\mathcal{E}})$. Industry efficiency becomes $\overline{\varepsilon}_{t+1} = e((\overline{x}_{t+1}, \overline{y}_{t+1})^{\mathcal{I}} + (\overline{x}_{t+1}, \overline{y}_{t+1})^{\mathcal{E}}; (x_{t+1}, y_{t+1})^{\mathcal{I}}, (x_{t+1}, y_{t+1})^{\mathcal{E}})$, where the first argument is equal to $(\overline{x}_{t+1}, \overline{y}_{t+1})$.

By Proposition 1, industry efficiency is less than the market share weighted harmonic mean of the firm efficiencies. It is straightforward to extend this aggregation result to groups of firms. In other words, industry efficiency is less than the market share weighted harmonic mean of the group of incumbent firms and the group of entrants. The difference is the inefficiency involved with the suboptimal balance between entrants and incumbents. Incumbents' efficiency is $\overline{\varepsilon}^{\mathcal{I}}_{t+1} = e((\overline{x}_{t+1}, \overline{y}_{t+1})^{\mathcal{I}}; (x_{t+1}, y_{t+1})^{\mathcal{I}}, (x_{t+1}, y_{t+1})^{\mathcal{E}})$ and entrants' efficiency is similar. Denote the market share for incumbents by $s^{\mathcal{I}}_{t+1} = \overline{p}_{t+1} y^{\mathcal{I}}_{t+1} / \overline{p}_{t+1} \overline{y}_{t+1}$, where the prices are determined by the year $t+1$ version of dual program (6), and similar for the entrants. Then $\overline{\varepsilon}_{t+1} \leq 1/(\frac{s^{\mathcal{I}}_{t+1}}{\overline{\varepsilon}^{\mathcal{I}}_{t+1}} + \frac{s^{\mathcal{E}}_{t+1}}{\overline{\varepsilon}^{\mathcal{E}}_{t+1}})$. This extension of Proposition 1 from firms to subclasses of firms suggests the following variations of Definition 1 and Proposition 6.

Definition 3. The *efficiency of entry*, $((x, y)^{\mathcal{I}}, (x, y)^{\mathcal{E}})$, equals $\varepsilon^E = \overline{\varepsilon}(\frac{s^{\mathcal{I}}}{\overline{\varepsilon}^{\mathcal{I}}} + \frac{s^{\mathcal{E}}}{\overline{\varepsilon}^{\mathcal{E}}})$, where $s^{\mathcal{I}}$ and $s^{\mathcal{E}}$ are the market shares of incumbents and entrants, respectively, evaluated at the prices determined by dual program (6).

Proposition 7. The Malmquist productivity index aggregates the efficiency of entry, incumbent and entrant efficiency changes, and technical change:

$$\overline{M}_t = \varepsilon^E_{t+1} \cdot \frac{\frac{1}{\overline{\varepsilon}_t}}{\frac{s^{\mathcal{I}}_{t+1}}{\overline{\varepsilon}^{\mathcal{I}}_{t+1}} + \frac{s^{\mathcal{E}}_{t+1}}{\overline{\varepsilon}^{\mathcal{E}}_{t+1}}} \cdot \sqrt{\frac{e((\overline{x}_t,\overline{y}_t),(x_t,y_t))}{e((\overline{x}_t,\overline{y}_t),(x_{t+1},y_{t+1}))} \cdot \frac{e((\overline{x}_{t+1},\overline{y}_{t+1}),(x_t,y_t))}{e((\overline{x}_{t+1},\overline{y}_{t+1}),(x_{t+1},y_{t+1}))}}$$

The first factor measures the dynamic industrial organization effect. Incumbent and entrant efficiency changes are aggregated in the second quotient by the market share weighted harmonic mean and market shares are evaluated at the shadow prices of the industry efficiency program (5). The square root measures technical change.

The middle factor accounts for the efficiency change of incumbent and entrants at an aggregated level, including not only firm efficiency changes but also the (static) industrial organization effect. This detail can be inserted as follows. Application of Proposition 3, $1/\overline{\varepsilon} = \frac{\sum s^i/\varepsilon^i}{\varepsilon^{IO}}$, to the incumbents and entrants transforms the middle factor in Proposition 7 to

$$\frac{\frac{1}{\overline{\varepsilon}_t}}{\frac{s^{\mathcal{I}}_{t+1}}{\overline{\varepsilon}^{\mathcal{I}}_{t+1}} + \frac{s^{\mathcal{E}}_{t+1}}{\overline{\varepsilon}^{\mathcal{E}}_{t+1}}} = \frac{\sum s^i_t/\varepsilon^i_t}{\varepsilon^{IO}_t} / \left(s^{\mathcal{I}}_{t+1} \frac{\sum_{\mathcal{I}} s^i_{t+1}/\varepsilon^i_{t+1}}{\varepsilon^{IO\mathcal{I}}_{t+1}} + s^{\mathcal{E}}_{t+1} \frac{\sum_{\mathcal{E}} s^i_{t+1}/\varepsilon^i_{t+1}}{\varepsilon^{IO\mathcal{E}}_{t+1}} \right).$$

This expression is a combination of firm efficiency changes, $\varepsilon^i_{t+1}/\varepsilon^i_t$, entrants efficiencies, ε^i_{t+1}, and industrial organization effects, $\varepsilon^{IO\mathcal{I}}_{t+1}/\varepsilon^{IO\mathcal{I}}_t$ and $\varepsilon^{IO\mathcal{E}}_{t+1}$.

Entry and exit is not the only departure from the basic model. The next one is markups. I defined an industrial organization by an allocation of inputs and outputs between the firms. All prices encountered in the analysis were endogenous, competitive prices. Even if the observed market prices are different, the competitive prices remain the right ones for the measurement of performance, just as in macroeconomics the Solow (1957) residual measures productivity growth if the value shares are based on competitive prices. To investigate the role of markups the observed prices must be included in the definition of an industrial organization and comparison with the shadow prices generated by the model yields the markups. Are markups good or bad for our performance indicator? The neoclassical answer is 'bad,' since they facilitate slack, and the Schumpeterian argument is 'good,' since they facilitate Research & Development funding. Not surprisingly, the evidence is not significant, but ten Raa and Mohnen (2008) have scratched the surface, decomposing the markups in capital and labor components. The first one is 'good' and the second not, suggesting that both the neoclassicals and the Schumpeterians have a point, but that the mechanisms differ (different factor markets).

Another, related departure of the basic model is to bring in increasing, decreasing or variable returns to scale. The latter is the counterpart of the U-shaped average cost curves and encompasses the former two. This variation drives a wedge between the technologies used in the firm and the industry efficiency programs, (1) and (5), which has to be specified (Färe and Grosskopf, 2004). Following Koopmans (1957) and Johansen (1972) variable returns to scale can be modeled by letting intensities sum to unity in the firm efficiency program and to the number of firms in the industry efficiency program. The number of firms in the industry efficiency program can be allowed to vary. The program becomes a mixed integer-linear program and, perhaps surprisingly, duality analysis can still be exploited to modify the results (ten Raa, 2009). While the relationship with the theory of contestable markets becomes less clean, the application to performance decomposition in firm terms and an industrial organization effect remains equally simple. The latter effect remains given by the ratio of Definition 1; the only modification remaining the inclusion of the adding-up constraints in the efficiency programs.

8 Conclusion

An industry may perform better, in the sense of productivity growth, by technical progress or by efficiency change. Both sources of growth have been decomposed in firm contributions, but the aggregation is known to be imperfect. The bias in the aggregation of the efficiencies of the firm reflects the allocative inefficiency in an industry. This industrial organization effect can be measured by the change in the ratio of the industry efficiency to the market share weighted harmonic mean of the firm efficiencies. The measure can be extended to capture the dynamic industrial organization effect of entry and exit.

References

[1] Balk, B.M. "The Residual: On Monitoring and Benchmarking Firms, Industries, and Economies with Respect to Productivity," *Journal of Productivity Analysis* 20, 1, 5-47 (2003)

[2] Bartelsman, E.J. and M. Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature* 38, 3, 569-94 (2000)

[3] Baumol, W.J. and D. Fischer, "Cost-Minimizing Number of Firms and Determination of Industry Structure," *Quarterly Journal of Economics*, 92, 3, 439-68 (1978)

[4] Baumol, W.J., J.C. Panzar and R.D. Willig, *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, New York (1982)

[5] Briec, W., B. Dervaux and H. Leleu, "Aggregation of Directional Distance Functions and Industrial Efficiency," *Journal of Economics* 79, 3, 237-61 (2003)

[6] Blackorby, C. and R.R. Russell, "Aggregation of Efficiency Indices," *Journal of Productivity Analysis* 12, 1, 5-20 (1999)

[7] Caves, D.W., L.R. Christensen, and W.E. Diewert, "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica* 50, 6, 1393-1414 (1982)

[8] Charnes, A., W.W.Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, 6, 429-44 (1978)

[9] Cherchye, L., C.A. Knox Lovell, W. Moesen, and T. Van Puyenbroeck, "One Market, One number? A Composite Indicator Assessment of EU Internal Market Dynamics," *European Economic Review* 51, 3, 749-79 (2007)

[10] Debreu, G., "The Coefficient of Resource Utilization," *Econometrica* 19, 3, 273-292 (1951)

[11] Denny, M., M. Fuss and Waverman L., "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries," in T.G. Cowing and R.E. Stevenson, eds., *Productivity Measurement in Regulated Industries*, Academic Press, New York (1981)

[12] Diewert, W.E., "The Measurement of Waste within the Production Sector of an Open Economy," *Scandinavian Journal of Economics* 85, 2, 159-79 (1983)

[13] Diewert, W.E., *Essays in Index Number Theory*, North-Holland, Amsterdam (1993)

[14] Diewert, W.E. "Benchmarking and the Nonparametric Approach to Performance Measurement," Tutorial presented at the University Autonoma of Barcelona (2005)

[15] Diewert, W. E., and K.J. Fox , "On Measuring the Contribution of Entering and Exiting Firms to Aggregate Productivity Growth," Discussion Paper No. 05-02, Department of Economics, University of British Columbia (2005)

[16] Färe, R., and S. Grosskopf, *New Directions: Efficiency and Productivity*, Kluwer Academic Publishers, Boston (2004)

[17] Färe, R., S. Grosskopf, B. Lindgren, and P. Roos, "Productivity Development in Swedish Hospitals: a Malmquist Output Index Approach," Discussion paper No. 89-3, Southern Illinois University, Illinois (1989)

[18] Farrell, M.J., "The Measurement of Production Efficiency," *Journal of the Royal Statistical Society* Series A, 120, 253-78 (1957)

[19] Førsund, F.R., and L. Hjalmarsson, "Generalised Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants," *Economic Journal* 89, 294-315 (1979)

[20] Fukuyama, H., and W.L. Weber, "Estimating Output Allocative Efficiency and Productivity Change: Application to Japanese Banks," *European Journal of Operational Research* 137, 177-190, 2002

[21] Georgescu-Roegen, N. "The Aggregate Linear Production Function and Its Applications to von Neumann's Economic Model," In: T. Koopmans, Editor, *Activity Analysis of Production and Allocation*, Wiley, New York, 98–115 (1951)

[22] Glass, J.C., G. McCallion, D.G. McKillop, and K. Stringer, "A 'Technically Level Playing-Field' Profit Efficiency Analysis of Enforced Competition Between Publicly Funded Institutions," *European Economic Review* 50, 6, 1601-26 (2006)

[23] Johansen, L., *Production Functions*, North-Holland, Amsterdam (1972)

[24] Jorgenson, D., M. Ho and K. Stiroh, "Growth of US Industries and Investments in Information Technology and Higher Education," *Economic Systems Research* 15, 3, 279-325 (2003)

[25] Koopmans, T.C., *Three Essays on the State of Economic Science*, McGraw-Hill, New York (1957)

[26] Sharkey, W.W., and L.G. Telser, "Supportable Cost Functions for the Multiproduct Firm," *Journal of Economic Theory* 18, 23-27 (1978)

[27] Shephard, R.W., *Theory of Cost and Production Functions*, Princeton University Press, Princeton (1970)

[28] Simar, L., and P.W. Wilson, "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes," *Journal of Econometrics* 136, 1, 31-64 (2007)

[29] Solow, R.M. "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics* 39, 3, 312-320 (1957)

[30] ten Raa, T., "The Substitution Theorem," *Journal of Economic Theory* 66, 2, 632-36 (1995)

[31] ten Raa, T., "Aggregation of Productivity Indices: The Allocative Efficiency Correction," *Journal of Productivity Analysis* 24, 2, 203-09 (2005)

[32] ten Raa, T., "Debreu's Coefficient of Resource Utilization, the Solow residual, and TFP: The Connection by Leontief Preferences," *Journal of Productivity Analysis* 30, 191-99 (2008)

[33] ten Raa, T., *The Economics of Benchmarking – Performance Measurement for Competitive Advantage*, Palgrave Macmillan, Houndmills and New York (2009)

[34] ten Raa, T., and P. Mohnen "Competition and Performance: The Different Roles of Capital and Labor," *Journal of Economic Behavior and Organization* 65, 3-4, 573-84 (2008)

[35] Tulkens, H., and P. Vanden Eeckaut, "Non-parametric Efficiency, Progress and Regress Measure for Panel Data: Methodological Aspects," *European Journal of Operational Research* 80, 474-499 (1995)

[36] Wolff, E.N. "Industrial Composition, Interindustry Effects, and the U.S. Productivity Slowdown," *The Review of Economics and Statistics* 67, 2, 268-77 (1985)