Benchmarking and industry performance

Thijs ten Raa

Abstract In this paper I interrelate productivity analysis and the theory of industrial organization. A proposition proves that an industrial organization is efficient if and only if it is supportable in the entry-proofness sense. Industrial performance is decomposed in efficiency and technical change terms as well as an industrial organization component. The performance measure is shown to be consistent with the Solow residual and Malmquist productivity indices for its components are provided.

Keywords Scope (dis)economies · Industrial organization · Efficiency · Aggregation · Productivity

JEL Classification L10 · D24 · O47

1 Introduction

In this paper I define benchmarking through a function that maps two arguments, namely the object to be benchmarked—a firm, represented by the pair of its input vector and its output vector—and the backdrop against which it is benchmarked—the industry—into a scalar, namely the efficiency of the firm. The mapping summarizes a program that identifies the best practice technologies or *benchmarks* for the firm. 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. Aggregate performance may rise more than firm performance scores suggest, if resources are better allocated between firms. The excess is the industrial organization effect.

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 in an activity analytic framework, but without using external price information (such as the world prices an open economy faces), as Nesterenko and Zelenyuk (2007) do. Another novelty is that my theory of benchmarking extends my interrelationship of two value-based performance measures—Diewert's (1981) 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.

There is more to the performance of a composite such as an industry than micro or firm performance and composition effects. Blackorby and Russell (1999) and Briec et al. (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. Indeed, Briec et al. (2003, Proposition 5) and Färe and Grosskopf (2004, p. 108) noticed a bias, ten Raa (2005) related it to the departure of the industrial efficiency condition that marginal productivities are equalized, and Balk (2009) shows that in a growth accounting setting with independent disaggregated price and quantity information differential price changes add a term to aggregate productivity change. I now proceed to interpret the bias as an industrial organization inefficiency measure.

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

T. ten Raa (✉)
Faculty of Economics and Business, Tilburg University,
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: tenRaa@UvT.nl

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 closest framework is that of Li and Ng (1995), who distinguish technical efficiency, allocatively efficiency and reallocative efficiency; the latter corresponds to my measure of industrial organization efficiency. They measure all efficiencies using the shadow prices of the efficiency program of the industry as a whole, while I employ the individual firm shadow prices for the micro efficiencies and show that departures between the two signal organizational inefficiency.

2 Benchmarking: price analysis

Denote firm i's input vector by $x^i \neq 0$ and its output vector by $y^i \neq 0, i = 1, \ldots, I$.[1] Input vectors have a common dimension, output vectors have a common dimension, and these two dimensions may differ (some positive integers). 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. An *industrial organization* can be represented by the allocation $(x^i, y^i)_{i=1,\ldots,I}$, 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 firm's actual output level and the best practice output level achievable with its 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 $j = 1, \ldots, I$, and to run the latter with intensities θ_j, as to inflate the output, y^i, by an expansion factor $1/\varepsilon$:[2]

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

Program (1) assumes that activities (x^i, y^i) can be run with *constant returns to scale* and that inputs and outputs

are *freely disposable*.[3] Let ε^i optimize primal program (1). The expanded y^i/ε^i is the *potential* output of firm i, using the best practice technologies. 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; these solve the dual program:[4]

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

This dual program is essentially the original tack to Data Envelopment Analysis, taken by Charnes et al. (1979).[5] 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 Eq. (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 and not necessarily shadow prices. Under constant returns to scale the ratio should equal unity for profit maximizers. A 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. For example, if the output mix of a firm is relatively intensive in terms of some input, the shadow price of that input will be high.

We now apply the apparatus to the efficiency of the industry. The definition goes back to Farrell (1957) and Førsund and Hjalmarsson (1979), who call it structural efficiency. Related, Johansen (1972) defines potential

---

[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).

[2] I do not surround the objective and constraints by the set symbol {.}. I omit $j = 1, \ldots, I$ under the summation model, following the Einstein summation convention.

---

[3] 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. Under constant returns to scale output and input based benchmarking procedures are equivalent. The program is linear in the nonnegative variables $1/\varepsilon, \theta_j$.

[4] 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 $\theta_j = 0, j \neq i$).

[5] Under the alternative scheme of input contraction, the dual would maximize the value of output.

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 $x = \sum x^i$, as to inflate the aggregate output vector, $y = \sum y^i$, by an expansion factor $1/\varepsilon$. Under constant returns to scale the industry technology is the cone spanned by the firm technologies, which is the same technology as the reference technology in the firm efficiency program (1). Hence the only modification to the latter is the replacement of firm resources and output target by industry resources x and output target y:

$$\max_{\varepsilon, \theta_j \geq 0} 1/\varepsilon : \sum \theta_j x^j \leq x, \quad y/\varepsilon \leq \sum \theta_j y^j \qquad (5)$$

Let ε 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 w and p, for the inputs and outputs, respectively; these *industry prices* solve the dual program:[6]

$$\min_{p', w' \geq 0} w'x : p'y^j \leq w'x^j, \quad p'y = 1, \text{ all } j \qquad (6)$$

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

$$1/\varepsilon = wx \qquad (7)$$

Analogous to Eq. (4), industry efficiency becomes:[7]

$$\varepsilon = py/wx \qquad (8)$$

3 Industrial organization: an efficiency measure

Extending Briec et al. (2003, Proposition 5), Färe and Zelenyuk (2003, formula 2.12) and Färe and Grosskopf (2004, p. 108), Proposition 1 establishes a one-sided relationship between the efficiencies of the firms and the efficiency of the industry.

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

Proof In the dual program (2), consider the socially optimal prices p/py^i and w/py^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), $(w/py^i)x^i \geq w^i x^i$ or $wx^i \geq py^i w^i x^i = py^i/\varepsilon^i$, using Eq. (3). Summing and invoking Eq.(7) and the price normalization constraint of (6), we obtain $1/\varepsilon = wx = w \sum x^i \geq \sum py^i/\varepsilon^i = \sum s^i/\varepsilon^i$. Inverting, industry efficiency becomes $\varepsilon \leq 1/\sum \frac{s^i}{\varepsilon^i}$. □

Extending Briec et al. (2003, p. 259), 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. 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.

Recall the formal model: Firms are represented by nonzero input-output activities $(x^i, y^i), i = 1, \ldots, I$, which can be run with constant returns to scale and free disposability. Given these data program (1) determines the firm efficiencies ε^i—with private shadow prices generated by dual program (2)—and program (5) determines industry efficiency ε—and the dual of the latter is used to define the market shares as in Proposition 1.

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 $\min_{p,w \geq 0} wx^i : py^j \leq wx^j, py = 1$, all j (normalizing against 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 (p, 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 $p^* y^j \leq w^* x^j, p^* y = 1$ and $w^* x < wx$. Hence $w^* x^i < wx^i$ for some i. This contradicts that (p, w) solves the rescaled dual program of that firm i. Since (p, 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 p/py^i and w/py^i are privately optimal. These private prices are collinear. □

Remark There may be corner/multiple solutions. For example, If two firms each produce one unit of output with one unit of labor and one unit of capital, then $(X, Y) = (x^i, y^i)_{i=1,2} = (((1, 1), 1), ((1, 1), 1))$ and the private input price vectors are (w_1, w_2) with nonnegative w_1, w_2 summing

---

to unity. The industry prices are the same, with the same multiplicity. When there is multiplicity, the relative price vectors are sets and Proposition 2 shows that if these sets are equal industry efficiency equals mean firm efficiency (and vice versa). I am grateful to a sharp referee for letting me complete this point.

The reason that industry efficiency is generally 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. This will be discussed after the defining industrial organization efficiency. 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 $\varepsilon^{\text{IO}} = \varepsilon \sum s^i / \varepsilon^i$, where s^i are the market shares evaluated at the prices determined by dual program (6).

Definition 1 frees Nesterenko and Zelenyuk (2007)'s definition of group revenue reallocative efficiency from its price dependence. 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. Bogetoft and Wang (2005) explore the potential efficiency gains of a better industrial organization, including what they call the harmony effect. Five simple examples show the measurability.

Examples

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

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. The shadow prices differ indeed. The shadow input price of firm 1 is (1, 2). (Under this price firms 1 and 3 break even.) The shadow input price of firm 2 is (2, 1).

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%.

5. Example 2 may also be varied as to have the price variation on the output side. Consider an industry that produces goods and services from a single input. Three firms each use one unit of input. Firm 1 produces one unit of the good, firm 2 produces one unit of the service, and firm 3 produces 2/3 units of either output. By the price normalization constraint in program (2), firm 1 has a good price of 1. The input price is 1. The service price must render the other two activities unprofitable, hence be in the interval [0, 1/2]. For firm 2 the output prices are opposite.

Analogous to Nesterenko and Zelenyuk's (2007, Proposition 1) revenue property, 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 et al. (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.

Definition 2 An industrial organization (X, Y) 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 et al. (1982) proposition that in a sustainable industry configuration total industry cost must be minimized. In Proposition 4 full industrial organization efficiency means $\varepsilon^{IO} = 1$ (Definition 1) and full firms efficiencies that the solutions to all programs (1) are $\varepsilon^i = 1$.

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^i solve 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 Eq. (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 (p, w) such that p/py^i and w/py^i are privately optimal. I claim that (p, w) supports the industrial organization. By Eq. (4), $wx^i = py^i$. 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. $\square$

4 Recovery of the Solow residual

Performance will be measured by the Solow (1957) residual, the difference between the output and input growth rates. Both growth rates are share-weighted expressions, using competitive valuations. In a noncompetitive environment the residual captures not only technical but also efficiency change. Since the latter concept is defined by the price-free benchmarking program (1), 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 $y_t^i, i = 1, \ldots, I$. 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:[8]

$$EC_t^i = \frac{d}{dt}\varepsilon_t^i / \varepsilon_t^i \qquad (9)$$

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). Program (1) determines the efficiency of firm i as a function of (x_t^i, y_t^i) and (X_t, Y_t), say $\varepsilon_t^i = e(x_t^i, y_t^i, X_t, Y_t)$. 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: $\varepsilon_t = e(x_t, y_t, X_t, Y_t)$. 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). Denote the row vectors of the two respective partial derivatives of

---

[8] The efficiency is the solution value to linear program (1), where the data, the inputs and outputs of the firms, appear in the constraint coefficients and bounds. The sensitivity with respect to the bounds is given by the shadow prices (Jansen et al. 1997). The differentiability need not be in the limited sense of functions but must be generalized (Rockafellar 1980). The sensitivity with respect to the constraint coefficients is differentiable (Freund 1985).

the mapping by e_1 and e_2. By total differentiation, the efficiency change of firm i is:

$$\frac{de}{dt}(x_t^i, y_t^i, X_t, Y_t)/e(x_t^i, y_t^i, X_t, Y_t)$$

$$= e_1(x_t^i, y_t^i, X_t, Y_t)\frac{d}{dt}(x_t^i, y_t^i)/e(x_t^i, y_t^i, X_t, Y_t) \quad (10)$$

$$+ e_2(x_t^i, y_t^i, X_t, Y_t)\frac{d}{dt}(X_t, Y_t)/e(x_t^i, y_t^i, X_t, Y_t)$$

The measurement of technical change involves a sign issue. 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* partial derivative with respect to the second argument—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:

$$TC_t^i = -e_2(x_t^i, y_t^i, X_t, Y_t)\frac{d}{dt}(X_t, Y_t)/e(x_t^i, y_t^i, X_t, Y_t) \quad (11)$$

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:

$$PG_t^i = e_1(x_t^i, y_t^i, X_t, Y_t)\frac{d}{dt}(x_t^i, y_t^i)/e(x_t^i, y_t^i, X_t, Y_t) \quad (12)$$

This definition of productivity growth is essentially the Solow residual, but because of its generality (the own effect of a firm on its efficiency, keeping the environment constant) this must be demonstrated. Now Solow (1957) defined his residual (applied to firm i) by $p_t^i\frac{d}{dt}y_t^i/p_t^i y_t^i - w_t^i\frac{d}{dt}x_t^i/w_t^i x_t^i$. His prices are implicit in the partial derivatives of my efficiency function. Proposition 5 makes it explicit, proving that expression (12) equals the Solow residual.

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 (12), 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 Eq. (4). $\square$

Summarizing, efficiency change is defined by (9), technical change by (11), productivity growth by (12), and the former two sum to the latter by Eq. (10), yielding Eq. (13):

$$PG_t^i = EC_t^i + TC_t^i \quad (13)$$

Things look only slightly different at the level of the industry. Now industry input and output, (x_t, y_t), are benchmarked against the frontier. The productivity growth of the industry is:

$$PG_t = e_1(x_t, y_t, X_t, Y_t)\frac{d}{dt}(x_t, y_t)/e(x_t, y_t, X_t, Y_t) \quad (14)$$

This expression is basically a summation of the firm productivity growth rates, (12), with the modification that private shadow prices have been replaced by social values.[9] The price difference constitutes a wedge which is precisely the aggregation bias uncovered by ten Raa (2005). I will now explain the role of industrial organization in the performance measure of productivity. Inspired by Caves et al. (1982), Färe et al. (1994) define the *Malmquist productivity index* by

$$M_t^i = \sqrt{\frac{e(x_{t+1}^i, y_{t+1}^i, X_t, Y_t)}{e(x_t^i, y_t^i, X_t, Y_t)} \cdot \frac{e(x_{t+1}^i, y_{t+1}^i, X_{t+1}, Y_{t+1})}{e(x_t^i, y_t^i, X_{t+1}, Y_{t+1})}}$$

$$= \frac{e(x_{t+1}^i, y_{t+1}^i, X_{t+1}, Y_{t+1})}{e(x_t^i, y_t^i, X_t, Y_t)}$$

$$\cdot \sqrt{\frac{e(x_t^i, y_t^i, X_t, Y_t)}{e(x_t^i, y_t^i, X_{t+1}, Y_{t+1})} \cdot \frac{e(x_{t+1}^i, y_{t+1}^i, X_t, Y_t)}{e(x_{t+1}^i, y_{t+1}^i, X_{t+1}, Y_{t+1})}}$$

$$\quad (15)$$

where the second line decomposes it in efficiency change and technical change. Similar to the firm index (15) the industry Malmquist productivity index is:

$$M_t = \sqrt{\frac{e(x_{t+1}, y_{t+1}, X_t, Y_t)}{e(x_t, y_t, X_t, Y_t)} \cdot \frac{e(x_{t+1}, y_{t+1}, X_{t+1}, Y_{t+1})}{e(x_t, y_t, X_{t+1}, Y_{t+1})}} \quad (16)$$

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

$$M_t = \frac{\varepsilon_{t+1}^{IO}}{\varepsilon_t^{IO}} \cdot \frac{\sum s_t^i/e(x_t^i, y_t^i, X_t, Y_t)}{\sum s_{t+1}^i/e(x_{t+1}^i, y_{t+1}^i, X_{t+1}, Y_{t+1})} \cdot$$

$$\sqrt{\frac{e(x_t, y_t, X_t, Y_t)}{e(x_t, y_t, X_{t+1}, Y_{t+1})} \cdot \frac{e(x_{t+1}, y_{t+1}, X_t, Y_t)}{e(x_{t+1}, y_{t+1}, X_{t+1}, Y_{t+1})}}$$

---

[9] Indeed, if the private and social prices would match, $PG_t = e_1(x_t, y_t, X_t, Y_t)\sum \frac{d}{dt}(x_t^i, y_t^i)/e(x_t, y_t, X_t, Y_t) = \sum PG_t^i \ e(x_t^i, y_t^i, X_t, Y_t)/e(x_t, y_t, X_t, Y_t)$.

Proposition 6 decomposes productivity growth in technical change and efficiency change and disaggregates the latter to the level of firms. This disaggregation involves a bias, which is accounted for by the industrial organization effect. 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 the second line of formula (15) to the industry and substitute, using Definition 1, $e(x_t, y_t, X_t, Y_t) = \varepsilon_t^{IO} / \sum s_t^i / e(x_t^i, y_t^i, X_t, Y_t)$ and similar for $e(x_{t+1}, y_{t+1}, X_t, Y_t)$. □

5 Illustration

Consider the Japanese banks ($i = 1, \ldots, I = 136$) over a five year period ($t = 1992, \ldots, 1996$).[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]

The last column of Table 1 shows that total productivity was initially stuck (0.08% growth in the period 1992–1993) but climbed strongly (to 3.15% in the period 1995–1996); on average it grew 1.17%. The bottom row decomposes this figure in technical change, efficiency change and the industrial organization effect. As usual, technical change is the most important source of productivity growth, contributing 0.85% per year. As is known from the growth accounting literature, technical change can be negative—see the second period. This means that the best practices were worse than in the year before. This phenomenon is

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

natural in mining, but less so in banking. Tulkens and Vanden Eeckaut (1995) redefine the frontier in terms of current *and* past practices; this technique would force technical change nonnegative and reduce efficiency change. Efficiency change can be negative—see the first period—but on average it augments technical change (0.85%) with a quarter (0.21%). Industrial organization is the third contributor to productivity growth, half the size of efficiency change. Thus, on average, the industrial organization of Japanese banking improved in the 1990s, but as for firms's efficiency it can go up and down.

This is a diagnosis of the Japanese banking industry. I do not explain the three components, but my hunch is that the advances in electronic banking drove technical change and the spread of ATMs enhanced efficiency. The industrial organization remained sluggish, leaving scope for competitive pressure, including changes in bankruptcy procedures.

6 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 reflects the inefficiency of the industrial organization. This paper offers measures for the latter.

---

[10] Because Malmquist productivity indices are ratios near 1 (a growth rate of 1% is an index of 1.01), we essentially have the *sum* of the three effects. Proposition 6 frees Zelenyuk (2006)'s analysis from its price dependence and the assumption of the Law of One Price.

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

References

Balk BM (2009) Measuring and relating aggregate and subaggregate productivity change without neoclassical assumptions. Discussion Paper 09026, Statistics Netherlands

Bartelsman EJ, Doms M (2000) Understanding productivity: lessons from longitudinal microdata. J Econ Lit 38(3):569–94

Baumol WJ, Fischer D (1978) Cost-minimizing number of firms and determination of industry structure. Quart J Econ 92(3):439–68

Baumol WJ, Panzar JC, Willig RD (1982) Contestable markets and the theory of industry structure. Harcourt Brace Jovanovich, New York

Blackorby C, Russell RR (1999) Aggregation of efficiency indices. J Prod Anal 12(1):5–20

Bogetoft P, Wang D (2005) Estimating the potential gains from mergers. J Prod Anal 23(2):145–71

Briec W, Dervaux B, Leleu H (2003) Aggregation of directional distance functions and industrial efficiency. J Econ 79(3):237–61

Caves DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output and productivity. Econometrica 50(6):1393–1414

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–44

Debreu G (1951) The coefficient of resource utilization. Econometrica 19(3):273–292

Diewert WE (1981) The economic theory of index numbers. In: Deaton A (eds) Essays in the theory and measurement of consumer behaviour in honour of Sir Richard Stone. Cambridge University Press, London, pp 163–208

Diewert WE (1983) The measurement of waste within the production sector of an open economy. Scand J Econ 85(2):159–79

Färe R, Grosskopf S (2004) New directions: efficiency and productivity. Kluwer Academic Publishers, Boston

Färe R, Zelenyuk V (2003) On aggregate Farrell efficiencies. Eur J Oper Res 146:615–20

Färe R, Grosskopf S, Lindgren B, Roos P (1994) Productivity development in Swedish hospitals: a Malmquist output index approach. In: Charnes A, Cooper W, Lewin AY, Seiford LM (eds) Data envelopment analysis: theory, methodology and applications. Kluwer Academic Publishers, Boston, pp 253–72

Farrell MJ (1957) The measurement of production efficiency. J R Stat Soc Ser A 120:253–78

Førsund FR, Hjalmarsson L (1979) Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. Econ J 89:294–315

Freund RM (1985) Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients. Math Program Study 24:1–13

Fukuyama H, Weber WL (2002) Estimating output allocative efficiency and productivity change: application to Japanese banks. Eur J Oper Res 137:177–190

Georgescu-Roegen N (1951) The aggregate linear production function and its applications to von Neumann's economic model. In: Koopmans T (eds) Activity analysis of production and allocation. Wiley, New York, pp 98–115

Jansen B, de Jong JJ, Roos C, Terlaky T (1997) Sensitivity analysis in linear programming: just be careful! Eur J Oper Res 101:15–28

Johansen L (1972) Production functions. North-Holland, Amsterdam

Koopmans TC (1957) Three essays on the state of economic science. McGraw-Hill, New York

Li SK, Ng YC (1995) Measuring the productive efficiency of a group of firms. Int Adv Econ Res 1(4):377–90

Nesterenko V, Zelenyuk V (2007) Measuring potential gains from reallocation of resources. J Prod Anal 28(1–2):107–16

Rockafellar RT (1980) Generalized directional derivatives and subgradients of nonconvex functions. Can J Math 32(2):257–80

Sharkey WW, Telser LG (1978) Supportable cost functions for the multiproduct firm. J Econ Theory 18:23–27

Shephard RW (1970) Theory of cost and production functions. Princeton University Press, Princeton

Solow RM (1957) Technical change and the aggregate production function. Rev Econ Stat 39(3):312–320

ten Raa T (2005) Aggregation of productivity indices: the allocative efficiency correction. J Prod Anal 24(2):203–09

ten Raa T (2008) Debreu's coefficient of resource utilization, the Solow residual, and TFP: the connection by Leontief preferences. J Prod Anal 30:191–99

Tulkens H, Vanden Eeckaut P (1995) Non-parametric efficiency, progress and regress measure for panel data: methodological aspects. Eur J Oper Res 80:474–499

Zelenyuk V (2006) Aggregation of Malmquist productivity indexes. Eur J Oper Res 174:1076–86