I.S.S.N: 1885-6888

ECONOMIC ANALYSIS WORKING PAPER SERIES

Malmquist Productivity Index Decompositions: A Unifying Framework

♦

José L. Zofío

Working Paper 12/2006

UNIVERSIDAD AUTONOMA DE MADRID

DEPARTAMENTO DE ANÁLISIS ECONÓMICO: TEORÍA ECONÓMICA E HISTORIA ECONÓMICA

Malmquist Productivity Index Decompositions: A Unifying Framework

José L. Zofío[a,*]

[a] *Departamento de Análisis Económico: Teoría Económica e Historia Económica. Universidad Autónoma de Madrid, E-28049 Cantoblanco, Madrid, Spain.*

June, 2004

Abstract

In two widely cited but unpublished working papers, Simar and Wilson (1998) and Zofío and Lovell (1998) proposed an alternative decomposition of the Malmquist Productivity Index, which retained what seemed to be the strongholds of previous proposals with regard to the contribution of technological and efficiency change to productivity change. Namely, a technical change term with regard to the best practice (VRS) technology which is to be found in Ray and Desli (1997) and a scale efficiency change term that illustrates a firm's situation with regard to optimal scale (benchmark technology), Färe, Grosskopf, Norris and Zhang (1994). Attaining this objective required the introduction of an additional term in the Malmquist Productivity Index decomposition, which would reflect the scale bias of technical change. It is our objective to provide economic rationale for this term within a theory of production context, the existing decompositions and recent articles that further elaborate on this issue. The ideas are illustrated using productivity trends in 17 OECD countries

Key Words: Productivity Change, Malmquist Indices, Distance Functions

JEL Codes: C43, D24, O47

[a] Tel: +34 914972406; fax: +34 914978616; e–mail: jose.zofio@uam.es

* This paper has benefited from conversations with C.A. Knox Lovell, Leopold Simar and Paul Wilson. The author is the sole responsible for any errors in the text. The earliest draft dates back to July 2001.

1. Introduction

Two decades ago Caves, Christensen and Diewert, CCD, (1982) theoretically introduced the now popular Malmquist index (MI) as the ratio between two distance functions that compares a firm's productivity with that of an alternative firm and, in a straightforward dynamic extension, over time. A decade later, Färe, Grosskopf, Lindgren and Roos, FGLR, (1989, 1994), in a working paper which dates back to 1989 showed how the MI could be empirically implemented by means of Data Envelopment Analysis, DEA, techniques, while proposing an initial decomposition. Drawing on the idea initially proposed by Nishimizu and Page (1982), these authors showed that in a Farrell (1957) context, productivity change based on Malmquist indexes can be decomposed into technological change and efficiency change, when allowing for inefficient production processes –*i.e.* a firm does not exploit the possibilities that the best practice frontier offers, but falls short from potential output.

However, by implicitly defining the MI with regard to what has been called a constant returns to scale-cone technology, the index imposes a technology representation that allows the comparison of a firm's productive performance to a maximum output to input ratio, a productivity ratio which is linked to the concept of returns to scale and scale efficiency, see Färe and Grosskopf (1998) or, more recently, Balk (2001). However, why imposing such technological restriction on the underlying technology when defining the Malmquist Index? In the original CCD (1982) Malmquist index this characterization of the technology was not present, culminating in an inaccurate measure of productivity change as Grifell-Tatjé and Lovell (1995) show, *i.e.* it ignores the contribution of scale change to productivity change.

This clearly called for a precise definition of what was to be understood as an "adequate" measure of productivity change. Researches soon agreed that extending the single input–single output ratio case to multiple variable production –where radial distance functions aggregate outputs and inputs, meant that the Malmquist index combining them had to fulfill several properties. Forsund (1997) summarizes this axiomatic approach to acknowledge an index as a *productivity* index, but the most relevant for the purpose at hand is the proportionality one. This property states that if outputs are to be increased in the same proportion from one period to the next while inputs remain the same, then the productivity index is to increase in the same proportion. Correspondingly, if inputs are reduced in the same proportion while outputs remain the same, then the productivity index should increase in such proportion. With regard to the specific Malmquist productivity indexes (MPI) this property requires that the distance functions which comprise it should be linearly homogeneous of degree +1 in outputs and –1 inputs, *i.e.* the benchmark technology characterizes by constant returns to scale.[1]

However, the fact that the supporting technology to correctly define productivity indexes corresponds to constant returns to scale does not mean that the underlying technology may not exhibit variable returns to scale. In fact, when identifying the contribution of returns to scale and scale efficiency one implicitly assumes that these terms have a role to play driving productivity change and, therefore, have to be included in the analysis. When doing so, two possibilities arise. Following Balk's (2001) terminology, 1) one may follow a top-bottom approach, decomposing the aggregate Malmquist productivity index initially proposed by FGLR (1989, 1994), which comply with the desirable proportionality property, but does not individualize the contribution that returns to scale and scale efficiency make to productivity change; 2) one can generalize the CCD (1982) index, which does not satisfy the proportionality property because it does not comprise the contribution that returns to scale and scale efficiency make to productivity change –but eventually satisfies it when scale change is included in the analytical formulation.

In the next section we introduce the necessary notation regarding technology and its distance functions representation. Section 3 summarizes the approaches followed by different authors trying to individualize the contribution of scale change to productivity change. This leads us to differentiate between the concepts of scale efficiency change and returns to scale and to show how they are interrelated. We make a distinction between these two concepts, as the existing literature clearly supports the idea that they are not interchangeable but complementary terms.

In section 4 we provide a meaningful theory of production interpretation of the scale bias of technical change, which can be considered as the link between the different decompositions proposed in the literature. The particular advantages and drawbacks of these proposals in uncovering and overlooking technological and efficiency change information are discussed in section 5. Here we focus on the theoretical work by Simar and Wilson (1998) and Zofío and Lovell (1998) –later applied but not justified by Wheelock and Wilson (1999)– to come up with a comprehensive decomposition of the MPI that would retain generally accepted definitions of these terms, while informing about the general framework where productivity change as well as technological and efficiency change take place –both from a technical and a scale perspective–. In this section we summarize the history surrounding the different

[1] Nevertheless, some authors believe that the axiomatic approach to index number theory, which relies on several desired properties to adequately define productivity indexes –*e.g.* proportionality, should not be strictly enforced: "At the risk of being labeled heretic, I see nothing "wrong" with estimating the Malmquist index based on empirical VRS technologies; we just need to be make sure that we and our readers are aware that it does not have an average product interpretation", Grosskopf (2003:465).

decompositions proposed in the literature and the contribution we make to ease their understanding. In this sense, our extended decomposition of the MPI, which has been extensively cited in the literature by Balk (2001), Ray (2001), Orea (2002), Lovell (2003) and Grosskopf (2003), falls into one of the most active research areas in productivity and efficiency measurement (see Olesen and Petersen, 2003). In section 6, we show how productivity change in OECD countries can be explained in the light of our extended decomposition.

Finally, we believe that this paper provides a meaningful interpretation of all the "building blocks" proposed in the literature to decompose the Malmquist productivity index, thus providing researches with an unifying framework where accurately interpret and choose among the existing decompositions.

2. Technology and Distance Functions

Consider a panel of i = 1,...,I producers observed in t = 1,...,T periods, transforming input vectors $x_i^t = (x_{1i}^t, \ldots, x_{Ni}^t) \in \Re^N_+$ into output vectors $y_i^t = (y_{1i}^t, \ldots, y_{Mi}^t) \in \Re^M_+$. Given these data, technology can be represented by the production possibility set of feasible input-output combinations:

$$S^t = \left\{ \left(x^t, y^t \right) : x^t \text{ can produce } y^t \right\}, \qquad t = 1, \ldots, T \tag{1}$$

which satisfies the usual Shephard (1970) or Färe and Primont (1995) axioms. Under this framework, a valid representation of the technology from the ith firm perspective is given by Shephard's output distance function[2].

$$D_{\mathrm{O}}^t\left(x_i^t, y_i^t\right) \equiv \inf_{\theta} \left\{ \theta > 0 : \left(x_i^t, y_i^t / \theta\right) \in S^t \right\}, \tag{2}$$

which is linearly homogenous of degree +1 in y and nonincreasing in x. If $D_{\mathrm{O}}^t\left(x_i^t, y_i^t\right)$ =1 the evaluated firm is said to be efficient belonging to the best practice technology –frontier– represented by the subset Isoq $S^t(x, y) = \{(x, y)$: $D_{\mathrm{O}}^t\left(x_i^t, y_i^t\right) = 1\}$. Therefore, if $D_{\mathrm{O}}^t\left(x_i^t, y_i^t\right) < 1$, a radial expansion of the output vector y_i^t is feasible within the production technology for the observed input level x_i^t, and the evaluated firm is said to be inefficient.

If period t technology were to exhibit global or local constant returns to scale, then the

[2] A complementary analysis could be developed from the input distance function perspective. However, using this orientation in what follows would not change any relevant issue regarding the decomposition of the Malmquist Productivity Index.

technology S^t implies a mapping $x \to y$ that is homogeneous of degree +1, i.e. $(x, y) \in S^t$ implies $(\lambda x, \lambda y) \in S^t$ for all $\lambda > 0$. This technology can be represented by

$$\breve{S}^t = \{(\lambda x^t, \lambda y^t) : (x^t, y^t) \in S^t, \lambda > 0\}. \tag{3}$$

The relevant consequence of this result is that the output distance function, if defined on a linearly homogeneous technology (3), is homogeneous of degree –1 in inputs –Färe and Primont (1995: 24), thus satisfying the condition that would render any Malmquist index based on a constant returns to scale technology a productivity index, see also Färe and Grosskopf (1996:54, proposition 3.2.6).

Clearly, whether the technology exhibits constant or variable returns to scale is to be determined with the sample data. However, if one assumes that the technology exhibits variable returns to scale, any Malmquist index based on the corresponding distance functions would not be regarded as a productivity index. Then, how can it be ensured that a Malmquist productivity index would satisfy the desirable homogeneity properties in outputs and inputs while retaining at the same time the variable returns to scale assumption on the technology? By defining distance functions that would compare productive performance to a benchmark linearly homogeneous technology which enhances such comparison from technical efficiency to include scale efficiency, *i.e.* which gauge productive efficiency. Balk (2001:16, eq. (16)) shows how this comparison corresponds to a distance function defined on the supporting –virtual– cone technology (3), which is equivalent to measure efficiency against firms operating at the most productive scale sizes, MPSSs, and whose productions processes characterize by local constant returns to scale. Thus, a distance function that encompasses technical and scale efficiency can be equivalently expressed as that one defined on the linear homogeneous extension (3) of the production possibility set (2). This distance function corresponds to

$$\breve{D}_{\mathrm{O}}^t(x_i^t, y_i^t) \equiv \inf_{\theta} \{\theta > 0 : (x_i^t, y_i^t / \theta) \in \breve{S}^t\} \tag{4}$$

$\breve{D}_{\mathrm{O}}^t(x_i^t, y_i^t)$ can be regarded as a measure of productive efficiency that compares a firm's observed productivity to the highest productivity level which corresponds to the highest scale elasticity. If $\breve{D}_{\mathrm{O}}^t(x_i^t, y_i^t) = 1$, then no productivity gains are feasible –either from a technical or a scale perspective. However, if $\breve{D}_{\mathrm{O}}^t(x_i^t, y_i^t) < 1$, the firm is productively inefficient and productivity gains can be achieved by increasing technical efficiency, scale efficiency, or both.

3. Decomposing the Malmquist index, MI

For any given firm i observed in two periods, t =1,2, (x_i^1, y_i^1) and (x_i^2, y_i^2), and using t =1 as benchmark technology, the original CCD (1982) Malmquist index defines as:

$$\mathrm{M}_{\mathrm{O}}^{1}\ (x_i^1, y_i^1, x_i^2, y_i^2) = \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)}, \tag{5}$$

where $D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)$ represents a mix period distance function which compares second period firms to the base period technology. In doing so, it is not mandatory that $\left(x_i^2, y_i^2\right) \in S^1$. In this case values of $D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right) > 1$ would be verified in the presence of technical progress, whose contribution to (5) can be singled out through the following decomposition:

$$\begin{aligned}\mathrm{M}_{\mathrm{O}}^{1}\ (x_i^1, y_i^1, x_i^2, y_i^2) &= \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} \cdot \frac{D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \\ &= \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) \cdot \mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)\end{aligned} \tag{6}$$

The Malmquist index (6) decomposes into a technical change and an efficiency change component. In an accepted interpretation of these terms, FGLR (1989,1994) stated that $\mathrm{TC}_{\mathrm{O}}^{1,2}\ (x_i^2, y_i^2)$ captures the shift in the technology between the two periods with regard to the actual best practice frontier, while $\mathrm{TEC}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ measures the change in relative efficiency, *i.e.* how far observed production is from maximum potential production. However, the index does not satisfy the proportionality property since it is not homogeneous of degree –1 in inputs. In the single input-single output case it does not measure productivity change understood as the change in average productivities as Grifell-Tatjé and Lovell (1995) show. Formally, this property requires that the Malmquist index (5) verifies $\mathrm{M}_{\mathrm{O}}^{1}\ (x_i^1, y_i^1, x_i^2, y_i^2) =$ $\mathrm{M}_{\mathrm{O}}^{1}\ (x_i^1, y_i^1, \mu x_i^2, \nu y_i^2) = \mu/\nu$.

Aware of this limitation, FGLR (1989, 1994) implicitly defined equal index but taking into consideration as benchmark technology not the actual best practice set (2) but its cone representation (3), which would render the Malmquist index (6) a productivity index, *i.e.*

$$\breve{\mathrm{M}}_{\mathrm{O}}^{1}\,(x_i^1, y_i^1, x_i^2, y_i^2) = \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \\ = \mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) \cdot \mathrm{EC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right), \tag{7}$$

while interpreting both technical and efficiency change terms in the same way. Unfortunately, the nature of what is measured completely changes as now the technical change term yields *potential* productivity change between firms that operate at the MPSSs –where firms are technical and scale efficient– in two consecutive periods –as argued before (4)–, *i.e.* $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ may be viewed as the highest productivity change in the absence of inefficiency[3]. Therefore $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ measures technical change with regard to the virtual supporting cone technology (3), and it would only correctly measure "technical change when constant returns to scale hold", Ray and Desli, RD, (1997: 1036). On the other hand, equal reasoning applies to the efficiency change term, which now measures how far a firm is from the benchmark cone productivity, and therefore comprises both technical and scale efficiency change terms –as Färe, Grosskopf, Norris and Zhang, FGNZ, (1994) would render later explicit in their enhanced and final decomposition.

Instead of working their way up from (6) to generalize the Malmquist index with a scale component that would take into account the contribution of returns to scale as proposed by Griffel-Tatjé and Lovell (1996, 1999), FGLR (1989, 1994) redefined the original Malmquist index into a productivity index by making use of the virtual cone technology (3). This forced them and later coauthors –FGNZ (1994)– to endorse the above interpretation of technical change which, nevertheless, does not correspond to the one commonly accepted –see the critics by RD (1997) and Balk (2001).

3.1. Interpreting technical efficiency change

Before further decomposing the productivity definition (7) of the Malmquist index (5), it is important to remark that the efficiency change term referred to the best practice technology in (6), $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, compares how a given firm varies its productive performance in time with regard to the base period technology –the Malmquist index– to how technology changes. Rearranging (5), one obtains

[3] Notice how potential technical change does not have to be led by a single producer, it is just the change between two periods productivity at optimal scale, which may be achieved by different producers in each period.

$$\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} = \frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} / \frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)} = \tag{8}$$
$$= \mathrm{M}_{\mathrm{O}}^1\ (x_i^1, y_i^1, x_i^2, y_i^2) / \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$$

From this perspective, increasing technical efficiency, $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)>1$, represents a final situation where the change in productivity outgrows technological change. If the latter outgrows the former, $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)<1$. Finally, when $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)=1$, the relative technical situation of the firm remains unchanged. This is depicted in figure 1, where the evaluated firm gets closer to the base period best practice frontier –captured by $\mathrm{M}_{\mathrm{O}}^1\ (x_i^1, y_i^1, x_i^2, y_i^2) > 1$, but contemporarily the production frontier experiences technical progress, $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)>1$. Since the productivity gain from increasing technical efficiency is exactly offset by technical progress, the distance from (x_i^t, y_i^t) to the best practice frontier is the same, $\mathrm{TE}_i^1 = y_i^1/\tilde{y}_i^1 = \mathrm{TE}_i^2 = y_i^2/\tilde{y}_i^2$, and there is no change in technical efficiency, *i.e.* $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{TE}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / \mathrm{TE}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) = \mathrm{M}_{\mathrm{O}}^1\ (x_i^1, y_i^1, x_i^2, y_i^2) / \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) = 1$.

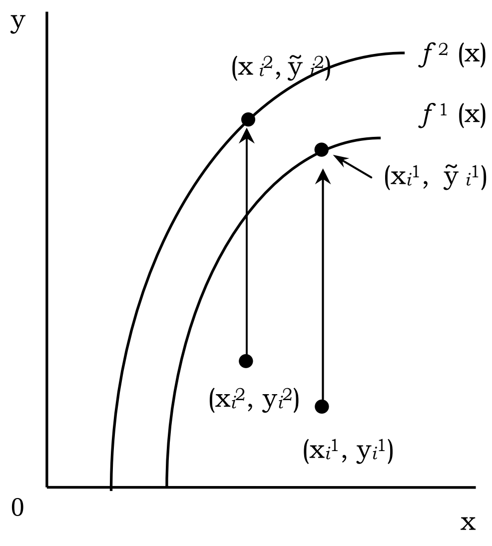

Figure 1. Interpreting Technical Efficiency Change

The Malmquist index (6) defines relative to the base period technology and technical change with regard to the firm observed in the comparison period, but it is possible to reverse this comparison structure. In this case,

$$\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)=\frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}=\frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)} / \frac{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}= \\ = \mathrm{M}_{\mathrm{O}}^2\ (x_i^1, y_i^1, x_i^2, y_i^2)/\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1\right) \tag{9}$$

Since both components will not generally yield the same result, one can define the geometric mean of both decompositions. Hence,

$$\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)=\frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}= \\ =\left[\frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} / \left[\frac{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)} \cdot \frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}\right]^{\frac{1}{2}} = \\ = \mathrm{M}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)/\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \tag{10}$$

3.2. Interpreting scale efficiency change

These results, based on the original CCD (1982) Malmquist index definition, just take into account technical information with regard to the best practice technology, but can be considered as benchmark for a parallel evaluation and interpretation of scale efficiency change.

Defining scale efficiency as any productivity differential due just to a suboptimal scale –i.e. the deviation from optimal scale that yields maximum productivity, MPSS–, and taking into consideration (2) and (4), one can derive a scale efficiency measure by means of the following index:

$$\mathrm{SE}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)=\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}, \tag{11}$$

If $D_{\mathrm{O}}^t\left(x_i^t, y_i^t\right)$ represents a technical efficiency measure which reflects how far is the evaluated firm from the best practice technology and $\breve{D}_{\mathrm{O}}^t\left(x_i^t, y_i^t\right)$ reflects how far it is from the highest productivity represented by the supporting –virtual– cone technology, then any difference between these two definitions corresponds to scale efficiency –since (4) represents both technical and scale efficiency while (3) only represents technical efficiency. Just as technical efficiency compares a firm’s productivity –actual output divided by its input level– to potential productivity in the best practice frontier –potential output divided by the input level, scale efficiency compares the highest –technically efficient– productivity attained at actual scale to the highest productivity observed at optimal scale.

In both cases, productivity differentials are assessed with respect to contemporary optima. If technical efficiency change (10) is the result of comparing technical efficiency in both periods, extending this concept to scale efficiency change requires the comparison of scale efficiency in both periods, i.e.

$$\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}. \tag{12}$$

If one agrees with this definition of scale efficiency change and the parallel process that leads to it departing from its technical efficiency change counterpart (10), it is possible to extend the idea of scale efficiency change as the final net result of comparing how a firm's changes its productive performance from a scale perspective to how technology's optimal scale changes. We consider that while moving from the base to the comparison period, a firm can improve its productive performance making use of the returns to scale offered by the best practice technology, while it is quite likely that at the same time the nature of the best practice technology with regard to optimal scale also changes from one period to the next.

These changes can be rendered explicit by decomposing scale efficiency change along the lines already introduced for the technical efficiency change case:[4]

$$\begin{aligned}
\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) &= \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} = \\
&= \frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \Big/ \frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)} = \\
&= \mathrm{RTS}_{\mathrm{O}}^1\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right),
\end{aligned} \tag{13}$$

where $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ represents productivity variations coming from a change in the scale

---

[4] It is interesting to note that both Balk (2001) and Lovell (2003) seem to be concerned by the fact that $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ does not only combine quantity vectors from both periods but also from both period technologies. However, the same claim could be extended to the previous technical efficiency change term $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, where both periods' quantity vectors and technologies are considered. Regarding $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ Balk (2001:172) concludes that "there seems to be to be some double-counting of technical change here" but there isn't because as presented in (13), $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ is the net result of comparing returns to scale to the scale –bias– of technical change and this last term is the only one to include both period technologies –just as $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in (9) is the net result between productivity change and technical change and it is in this last term where both technologies can be found. Lovell (2003) states that $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ "must combine the effects of scale economies and technical change". This is exactly what is presented in (13) and discussed in what follows, *i.e.* $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ combines the effect of returns to scale and the scale –bias– of technical change.

of the evaluated firm with respect to the base technology, *i.e.* returns to scale –more on this in the following section, while $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ represents productivity variations on scale efficiency coming from the change in the technology with regard to the comparison period firm, *i.e.* the scale –bias– of technical change. If one takes into account the second period technology to measure returns to scale and the base period firm to measure the scale –bias– of technical change, it is possible to express scale efficiency change as the geometric mean of these two indexes:

$$
\begin{aligned}
\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) &= \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} = \\
&= \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} / \\
&/ \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}\right]^{\frac{1}{2}} = \\
&= \mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right).
\end{aligned} \tag{14}
$$

4. Interpreting returns to scale and the scale –bias– of technical change

4.1 Returns to scale

The different components in which scale efficiency change can be decomposed refer to several terms already proposed in the Malmquist index literature. The second line of eq. (14), $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, corresponds to what RD (1997) initially referred to as scale efficiency change, as well as Grifell-Tatjé and Lovell (1996, 1999) and Balk (2001)[5]. However, this term clearly differs from the one introduced in (12), as the latter uses a single period technology while scale efficiency change compares scale efficiency with regard to own period technologies, i.e. how the firm moves toward or away from optimal scale in both periods. In an interpretation that illustrates the nature of this term, Orea (2002) and Lovell (2003) make use of discrete time formulations that identify it as a measure of the contribution of returns to scale to productivity change.

[5] Grifell-Tatjé and Lovell (1996, 1999) propose $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^1\right)$ as the scale effect index, which is equivalent to the one presented in (14) since from the output perspective, it is homogeneous of degree 0 in outputs as long as $y^2=\lambda y^1$, $\lambda>0$, making irrelevant which output level, y^1 or y^2, is chosen. Balk (2001) and Lovell (2003) explicitly consider the contribution to productivity change of any change in the output mix, i.e. $y^2\neq\lambda y^1$, $\lambda>0$. However, as it is not the scope of this paper, we assume that this term plays no role, so the subsequent proposals made by Grifell-Tatjé and Lovell (1996, 1999), Balk (2001) and Lovell (2003)

On the basis of a translog output-oriented parametric definition of productivity change, which includes technical change, technical efficiency change and a remaining scale effect, both authors defend that the latter term corresponds to $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in (14). Orea (2002: 9, eq.(4)) states that $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ measures the "contribution of scale economies to productivity growth indirectly, that is, through comparisons … with the most productive scale size". This author continues by providing an alternative way to assess the contribution of scale economies "without any reference to scale efficiency" *ibid.* pag. 12, implicitly stating that $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ must not be conceptually mixed up with $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. From the same perspective, Lovell (2003) supports this interpretation for $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, stating that it "provides a valid measure of the contribution of scale economies". His next statement implicitly supports the distinction we are trying to make here between scale efficiency change, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, and returns to scale, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$: "An important implication … is that change in scale efficiency plays no explicit role in the decomposition of the Malmquist productivity index". Finally, Ray (2001) also seems to acknowledge some difficulties when interpreting $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ as a measure of scale efficiency change, stating that this term "is less easy to interpret". Nevertheless he keeps addressing it in such way "this (term) can be called the scale (efficiency) factor", but denotes it by SCF (scale change factor) in order to differentiate it from $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ –as proposed by FGNZ (1994)[6].

To reinforce the interpretation of the second line in (13) as an index which measures the contribution of returns to scale to productivity change, let us consider the next alternative decomposition of the Malmquist productivity index (7):

$$
\begin{aligned}
\breve{\mathrm{M}}_{\mathrm{O}}^{1}\,(x_i^1, y_i^1, x_i^2, y_i^2) &= \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)} = \\
&= \mathrm{M}_{\mathrm{O}}^{1}\,(x_i^1, y_i^1, x_i^2, y_i^2) \cdot \mathrm{RTS}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1, x_i^2, y_i^2\right),
\end{aligned}
\tag{15}
$$

which is the original CCD (1982) Malmquist index enhanced with the contribution of returns to scale to productivity change. Grifell-Tatjé and Lovell (1996, 1999:85), were the first authors to propose (15) as a way to generalize the original index with a term which would take into

coincide with the one initially introduced by RD (1997).

[6] It is interesting to note that the translog parametric definition of scale efficiency, which goes back to Ray (1998) in the single output case and Balk (2001) in the multiple output case, is not disputed here. We just support the alternative scale efficiency *change* term, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in (12), without giving up the information provided by $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, which nevertheless should not be named scale

account the contribution of scale economies to productivity change. Even if they kept terming the second component of the right hand side of (15) as scale efficiency change –in opposition to (12), its meaning was the one just stated. In fact $\tilde{M}^1_O(x^1_i, y^1_i, x^2_i, y^2_i) \gtrless M^1_O(x^1_i, y^1_i, x^2_i, y^2_i)$ "depending upon the local … nature of scale economies characterizing period *t* technology … Locally increasing (decreasing) returns to scale produces an upward (downward) adjustment to the conventional Malmquist productivity index", *ibid.* pag. 86.

Thus, If $\mathrm{RTS}^{1,2}_O\left(x^1_i, y^1_i, x^2_i, y^2_i\right) > 1$, the firm improves its performance on a scale basis with regard to the base period productivity benchmark by exploiting increasing returns to scale and getting closer to the MPSS. Contrarily, $\mathrm{RTS}^{1,2}_O\left(x^1_i, y^1_i, x^2_i, y^2_i\right) < 1$ indicates that input change carries decreasing returns to scale and the firm is moving away from optimal scale. Finally, when $\mathrm{RTS}^{1,2}_O\left(x^1_i, y^1_i, x^2_i, y^2_i\right) = 1$, the firm does not profit (endure) from scale economies (diseconomies) as when constant returns to scale prevail over the input range $\left[x^1_i, x^2_i\right]$. Figure 2 illustrates $\mathrm{RTS}^{1,2}_O\left(x^1_i, y^1_i, x^2_i, y^2_i\right) > 1$ when considering the first period technology as benchmark. Here the firm (x_i^1, y_i^1) profits from increasing returns to scale when moving toward (x_i^2, y_i^2) –increasing average production along the best practice frontier $f^1(x)$ from the efficient projection $(x_i^1, \tilde{y}_i^1)$ toward the MPSS represented by x_i^{1*}, *i.e.* it becomes scale efficient from the base period perspective.

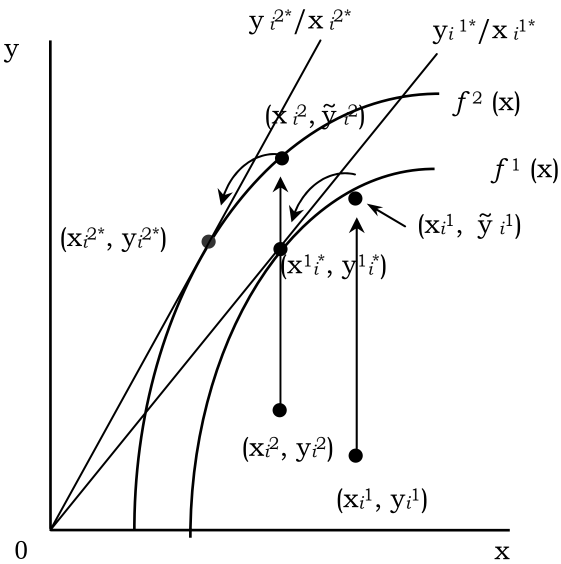

Figure 2. Interpreting Scale Efficiency Change

efficiency change if one wants to avoid conflicting denominations.

However, it is quite likely that while the evaluated firm gets closer to the base period optimal scale represented by (x_i^{1*}, y_i^{1*}), this optimal scale contemporarily changes, rendering such attempt to improve its scale performance useless. This is what happens in figure 2, where optimal scale moves from (x_i^{1*}, y_i^{1*}) to (x_i^{2*}, y_i^{2*}). Hence, the productivity differential due to the inefficient scale of (x_i^t,y_i^t) with regard to the highest productivity experienced at optimal scale is the same in both periods, $\mathrm{SE}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)$ = ($\tilde{y}_i^{1}/x_i^1$) / (y_i^{1*}/x_i^{1*}) and $\mathrm{SE}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)$= ($\tilde{y}_i^{2}/x_i^2$) / ($y_i^{2*}/x_i^{2*}$). As a result there is no change in scale efficiency, *i.e.* $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = $\mathrm{SE}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)$ / $\mathrm{SE}_{\mathrm{O}}^{1}\left(x_i^1, y_i^1\right)$ = $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ / $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = 1. In this case, the productivity increase obtained by (x_i^1, y_i^1) by reducing its productive scale toward the base period optimal scale, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$ –experimenting increasing returns to scale, is exactly offset by a contemporary reduction in optimal scale from x_i^{1*} to x_i^{2*} which leaves the evaluated firm in an scale inefficient position.

4.2 Scale –bias– of technical change

This result is captured by the third line in (14), which has been termed by Simar and Wilson (1998: 9-10) and Zofio and Lovell (1998:4) as the scale –bias– of technical change, $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, because it is the scale counterpart of $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Just as $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ represents the benchmark to assess if any technical gain of the firm finally results in a technical efficiency gain when moving toward the best practice frontier –eq. (10), $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ plays the same role by showing whether any productivity gain (loss) due to the effect of increasing (decreasing) returns to scale with respect to the benchmark technology, finally results in a scale efficiency gain or not –eq. (14). This can be emphasized by jointly taking into account technical change and the scale –bias– of technical change to determine potential productivity change over time from a given firm perspective, *i.e.* productivity change at the reference optimal scale. In this scheme, it is possible to recall this term –introduced in (7)– and decompose it in the following way:

$$
\begin{aligned}
\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) &= \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} = \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} = \\
&= \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) = \\
&= \frac{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right) / \breve{D}_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^{1}\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^{2}\left(x_i^2, y_i^2\right)} = \\
&= \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) \cdot \breve{\mathrm{T}}\mathrm{C}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) / \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right).
\end{aligned} \tag{16}
$$

From the firm's comparison period perspective, potential productivity change at optimal scale can be decomposed into technical change –first term in the right hand side of (16)– weighted by a bias against or in favor of the reference firm input scale. This can be easily shown rearranging $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ as in the third line of (16). The numerator corresponds to productivity change at optimal scale while the denominator corresponds to productivity change coming from technical change at the reference input scale, i.e. $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ = $\breve{\mathrm{T}}\mathrm{C}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) / \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$.

If $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) > 1$, productivity gain reflected by technical change at the comparison period input scale does not match the potential productivity change observed at optimal scale, and accordingly, technical change at the firms' scale has to be augmented with an additional productivity gain if it is to match that one at optimal scale. Therefore, we can conclude that the change in the technology with regard to optimal scale presents a bias against the reference input scale, which would be the interpretation for $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right)$ when expressed as in the first line of (16). Contrarily, when $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) < 1$, productivity change at the reference input scale exceeds productivity change at optimal scale, and consequently technical change has to be lowered in the amount necessary to match productivity change at optimal scale. Therefore, the change in the technology with regard to optimal scale presents a bias in favor of the reference input scale. Finally, $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^2, y_i^2\right) = 1$ shows how the scale bias of technical change is neutral since productivity change at the reference input scale matches productivity change at optimal scale, as would be the case in the presence of constant returns to scale[7].

Let us now interpret the alternative values of the scale bias of technical change with respect to returns to scale and their net result regarding scale efficiency change –eq. (14). From a geometric mean perspective, if $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$, the scale bias of technical change works against the reference input scales –as in eq. (16)– and any productivity gain due to increasing returns to scale, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$ would be counterbalanced. Therefore, whether there is scale efficiency gain or not will depend on $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \underset{<}{\geq}$ $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. On the other hand, if the firm undergoes decreasing returns to scale from the base to the comparison period, productivity loss is reinforced and the firm losses scale

[7] Zofio and Lovell (1998) state that if $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ equals unity "technical change is neutral with respect to scale because it has not altered the technically optimal scale". However, Ray (2001) shows how this numerical outcome is also compatible with technological changes where optimal scale changes. All it is necessary is that productivity change at the reference input scale matches productivity change at optimal scale.

efficiency, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ / $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ < 1. When $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ < 1, the scale bias of technical change works in favor of the reference input scale. Hence, if the firm experiences decreasing returns to scale when moving from the first to the second period, this productivity loss would be offset by the scale bias of technical change and the final result on scale efficiency change once again depends on their relative values. If the firm enjoys increasing returns to scale, the scale bias of technical change reinforces such productivity gain and $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ / $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ < 1. Finally if $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = 1, technical change at the reference and optimal inputs scales coincide, and any change in scale efficiency is exclusively given by the nature of returns to scale as the scale bias of technical change is neutral.

5. Decomposing the Malmquist productivity index, MPI

The key question regarding the above developments is whether it is possible to propose a Malmquist productivity index decomposition that provides all relevant information regarding technological and efficiency change, and whose terms can interpreted in an meaningful manner. One way to proceed is to chronologically assess the relative advantages and drawbacks of the alternative decompositions proposed in the literature.

The initial and still most popular decomposition of the MPI is the one proposed by FGNZ (1994), which enhances the one presented in (7) to take into account a scale component. Considering its geometric mean definition, it is equal to

$$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} =$$

$$= \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} \cdot \frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right) / D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right) / D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} = \tag{17}$$

$$= \mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right).$$

This decomposition stresses a definition of technical change $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ which corresponds to potential productivity change at optimal scale –the shift in the virtual supporting cone technology–, but overlooks the change in the best practice technology, *i.e.* the usual definition of technical change $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. On the other hand, it informs about technical efficiency change $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ and scale efficiency change $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, defined as argued in the previous section. Clearly, technical change measured on the benchmark cone technology neglects the shift on the best practice frontier and may overstate or underestimate this latter value. "Hence, the [FGNZ (1994)] technical change component must include something else", Lovell (2003). This something else is the scale bias of technical change as presented and discussed in eq. (16) [8].

Following in time is the RD (1997) proposal, which coincides in the single output case with that of Grifell-Tatjé and Lovell (1996,1999), Balk (2001) and Lovell (2003). This decomposition corresponds to

$$
\begin{aligned}
\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\,(x_i^1, y_i^1, x_i^2, y_i^2) &= \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}\cdot\frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} = \\
&= \left[\frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}\cdot\frac{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}}\cdot\frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}\cdot \\
&\quad\cdot\left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)/D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)/D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}\cdot\frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)/D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)/D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right] = \\
&= \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)\cdot\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)\cdot\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)
\end{aligned}
\tag{18}
$$

This proposal would be the most widely accepted one if we were to take into account the number of academics who endorse it. It measures the contribution of best practice technical change $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, the undisputed factor representing technical efficiency change $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ and returns to scale, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$.

If we were to balance both proposals, it is interesting to highlight what these authors say in favor of and against each decomposition. With respect to (17), RD (1997), Grifell-Tatjé and Lovell (1999) and Balk (2001) note that $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ would not correctly measure technical change in the presence of variable returns to scale; productivity change at optimal scale is nothing but the potential productivity change that a firm could enjoy if it were producing efficiently from a technical and a scale perspective in both periods.

---

[8] Grifell-Tatjé and Lovell (1996, 1999:92) show how an upward bias in $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ mismeasures technical change by an amount that "creates a proportionally large downward bias in the FGNZ scale effect", This amount is precisely $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Since the relevant scale effect for this authors measures the contribution of returns to scale: $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, if $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ –an upward bias in technical change–, then $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$ and $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) < \mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ –the equivalent downward

$\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ is identified as a scale effect which does not attract much criticism –it is easy to interpret in the way already discussed–, but it is to be replaced by $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ if one wants to correctly assess technical change[9]. Substituting $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ for $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ does not pose a problem, as the former is also easy to interpret as a scale effect that takes into account the contribution of returns to scale[10]. On the other hand, Färe, Grosskopf and Norris (1997:1.042) exemplify how $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in (18) cannot be interpreted as a scale efficiency change component as it "may incorrectly identify the scale properties of the underlying technology". In fact, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ "contains no mix-period terms", which is what renders $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ unsuitable for scale efficiency change evaluations –as previously argued for a correct interpretation of technical efficiency change $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in both (17) and (18). Therefore, there is a trade off between technical change and scale efficiency change. If one supports a decomposition that includes the accepted notion of effective technical change at the firms input scale –and not potential productivity change at optimal scale–, then one gives up a scale efficiency change term but takes in a returns to scale component.

At this point, asking for an economically meaningful decomposition of the Malmquist productivity index is equivalent to discard any proposal whose terms cannot be interpreted in a theory of production context. However, both (17) and (18) decompose in terms which have a clear interpretation. There are a number of "building blocks" that can be combined in different but intelligible ways to produce the same MPI result. Therefore, if one were to reject one particular proposal, it would be on the grounds that some of its components cannot be interpreted in the way they claim. Nevertheless, they can be interpreted in the way already discussed. Therefore, besides cross criticisms, our conclusion is that all terms in which the alternative decompositions break down can be interpreted in a valid way. Regarding (18), it provides an accurate decomposition of productivity change taking into account firm's input scale for measuring both technical change and returns to scale. In (17), this desirable

---

bias in scale efficiency change with respect to the relevant returns to scale term–.

[9] Only Balk (2001) does not comment on $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. This author dismisses the interpretation of scale efficiency change as argued in (14), *i.e.* as the relative relationship between returns to scale and the scale –bias– of technical change which informs about the final situation with regard to optimal scale in both periods: $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ = $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ / $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. However, this provides rationale for a meaningful interpretation of these terms –including the scale –bias– of technical change beyond "a ratio of scale efficiencies" *ibid.* pag. 172.

[10] Leading Lovell (2003) to conclude that (18) "jettisons the notion of change in scale efficiency $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, a notion that I believe has misled researchers for years".

relationship between scale and productivity change coming from technical change is lost, but additional information regarding technological and efficiency changes is given, *i.e.* potential productivity change and scale efficiency change are now explicitly considered.

However, even if by choosing one of the two decompositions one has to sacrifice some information regarding technical and scale changes, both proposals are interrelated. In fact, from (16) $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ and from (14) $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Therefore, the scale –bias– of technical change $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ represents the cornerstone that links both decompositions, rendering possible a complete characterization of productivity change both from a technological –best practice– and efficiency perspective. Including $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ in the Malmquist productivity index decomposition would allow immediate access to all components that have been proposed in the literature. The question is whether it is possible to suggest a decomposition which includes the scale –bias– of technical change.

Being aware of the debate surrounding (17) and (18), Simar and Wilson (1998) and Zofío and Lovell (1998) introduced such decomposition. Their proposal can be obtained from both formulations. One may replace the potential contribution of productivity change at optimal scale in (17) by that of the effective contribution of technical change –productivity change of the benchmark technology at the firm's input scale– weighted by the scale –bias– of technical change –how productivity change at optimal scale shows a bias against or in favor of the firm's input scale, *i.e.* $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Alternatively, one may replace the effective contribution of returns to scale –how a firm profits from local increasing returns or endures local decreasing returns that materialize in higher or lower productivity change– by their counterpart in the form of the effective contribution of scale efficiency change –the movement of the firm toward or away from technically optimal scale in both periods– weighted by the scale –bias– of technical change, *i.e.* $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Proceeding in either way, one obtains the following decomposition:

$$
\begin{aligned}
\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\,(x_i^1, y_i^1, x_i^2, y_i^2) &= \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} = \\
&= \left[\frac{D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)} \cdot \frac{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)}\right]^{\frac{1}{2}} \cdot \frac{D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \\
&\cdot \frac{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)/ D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)/ D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)} \cdot \\
&\cdot \left[\frac{\breve{D}_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)/ D_{\mathrm{O}}^1\left(x_i^1, y_i^1\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)/ D_{\mathrm{O}}^2\left(x_i^1, y_i^1\right)} \cdot \frac{\breve{D}_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)/ D_{\mathrm{O}}^1\left(x_i^2, y_i^2\right)}{\breve{D}_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)/ D_{\mathrm{O}}^2\left(x_i^2, y_i^2\right)}\right]^{1/2} \\
&= \mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \\
&\cdot \mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)
\end{aligned} \tag{19}
$$

All these terms have been previously interpreted, but given the number of scholars who advocate using (18), it is important to remark that the contribution of returns to scale is implicitly considered in (19) through (14). By jointly looking at scale efficiency change and the scale –bias– of technical change, we can obtain relevant information with regard to returns to scale. Rephrasing the discussion in section 4.1, if the firm gains scale efficiency from the base to the comparison period, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$, while the scale bias of technical change works against the firm's reference input scale $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$, this outcome is only possible if returns to scale make a positive contribution to productivity change, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$ –a contribution which is larger than the unfavorable change in the scale bias of technical change. On the other hand, if a scale efficiency gain is accompanied by a favorable scale change of the technology, $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) < 1$, then the presence of increasing returns to scale reinforces such scale efficiency gains. Alternatively, if decreasing returns to scale reduce productivity change, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) < 1$, then scale efficiency gains are still possible as long as the favorable scale –bias– of technical change is not counterbalanced by those lowering returns, $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ –where both terms are smaller than one. In both cases the final outcome would be $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) > 1$ and an opposite discussion may be presented when scale efficiency change reduces $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) = \mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) / \mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) < 1$.

As Simar and Wilson (1998: 11) remark, all this information would be lost if one settles for (18), because one would know the contribution of returns to scale to productivity change, but would not know if such contribution finally results in scale efficiency gain or not, $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \gtreqless 1$, neither if the change experienced by optimal scale works against or in favor of the firm's reference scale, $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right) \gtreqless 1$.

5.1 Summarizing the history of MPI decompositions.

The different parts of the MPI decomposition puzzle are presented in Table 1 as they were introduced in the literature. Here, the initial Caves *et al.* (1982) index $\mathrm{M}_{\mathrm{O}}^{1,2}\,(x_i^1, y_i^1, x_i^2, y_i^2)$ does not comply with the proportionality property, which derives from ignoring the impact of returns to scale on productivity change, and therefore it does not constitute a productivity index. In order to define a MPI definition that would comply with such property, Färe et al. (1989,1994) followed a top-down approach which yielded an index which measures productivity $\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\,(x_i^1, y_i^1, x_i^2, y_i^2)$ by implicitly incorporating the effect of returns to scale $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ –proposing an initial decomposition into potential productivity gain $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ and efficiency change $\mathrm{EC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$. Unfortunately, when trying to individualize the scale contribution, FGNZ (1994) endorsed the technical change component inherited from FGLR (1989, 1994) –which corresponds to productivity change at optimal scale, believing that the contribution of scale change was adequately identified by decomposing efficiency change into (pure) technical efficiency change $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ and scale efficiency change $\mathrm{SEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$.

Table 1. Alternative Malmquist Productivity Index decompositions

	Product. Change	Potential Product. Change (1)	Eff. Change(2)	Tech. Change	Tech. Eff. Change	Returns to Scale (3)	Scale Eff. Change	Scale of Tech. Change
Proposal	$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}$	$\mathrm{PTC}_{\mathrm{O}}^{1,2}$	$\mathrm{EC}_{\mathrm{O}}^{1,2}$	$\mathrm{TC}_{\mathrm{O}}^{1,2}$	$\mathrm{TEC}_{\mathrm{O}}^{1,2}$	$\mathrm{RTS}_{\mathrm{O}}^{1,2}$	$\mathrm{SEC}_{\mathrm{O}}^{1,2}$	$\mathrm{STC}_{\mathrm{O}}^{1,2}$
$\mathrm{M}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ CCD (1982)	NO	NO	NO	YES	YES	NO	NO	NO
$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ FGLR (1989,1994)	YES	YES	YES	NO	NO	NO	NO	NO
$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ FGNZ (1994)	YES	YES	YES	NO	YES	NO	YES	NO
$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ RD (1997) (4)	YES	NO	NO	YES	YES	YES	NO	NO
$\breve{\mathrm{M}}_{\mathrm{O}}^{1,2}\ (x_i^1, y_i^1, x_i^2, y_i^2)$ SWZ (1998)	YES	YES	YES	YES	YES	YES	YES	YES

(1) $\mathrm{PTC}_{\mathrm{O}}^{1,2} = \mathrm{TC}_{\mathrm{O}}^{1,2} \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}$

(2) $\mathrm{EC}_{\mathrm{O}}^{1,2} = \mathrm{TEC}_{\mathrm{O}}^{1,2} \cdot \mathrm{SEC}_{\mathrm{O}}^{1,2}$

(3) $\mathrm{RTS}_{\mathrm{O}}^{1,2} = \mathrm{SEC}_{\mathrm{O}}^{1,2} \cdot \mathrm{STC}_{\mathrm{O}}^{1,2}$

(4) Ray and Desli (1997) proposal is equivalent to that of Grifell-Tatjé and Lovell (1996, 1999), Balk (2001) and Lovell (2003) in the single output case.

Unconvinced by the existing definition of technical change, Grifell-Tatjé and Lovell (1996,1999) followed a bottom-up approach departing from the initial CCD (1982) definition, which coincides with the RD (1997) proposal –who, nevertheless, followed a top-down approach from FGNZ (1994) MPI definition, but rejected their decomposition. Both sets of authors identified the commonly accepted definition of technical change, $\mathrm{TC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, adopted the technical efficiency change component, $\mathrm{TEC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, and claimed a different definition of scale efficiency change which is inconsistent with the parallel notion of technical efficiency change –a fact that was criticized by Färe, Grosskopf and Norris (1997). Grifell-Tatjé and Lovell (1996,1999:88) clearly suggested that the new scale efficiency term, even if called in such way, really captures the contribution of returns to scale. The work by Orea (2002) and Lovell (2003) provide further rationale for supporting this interpretation, and so it can be identified with $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$.

However, even the proposal made by RD (1997) and its equivalent Grifell-Tatjé (1996,1999), Balk (2001) and Lovell (2003) counterparts identify the role of scale when defining technical change and the contribution of returns to scale, it disregards productivity change at optimal scale and the change in scale efficiency. Simar and Wilson (1998) and Zofio and Lovell (1998) uncovered the concept of scale –bias– of technical change $\mathrm{STC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ offering a decomposition whose terms could be interpreted within the theory of production context, and whose combination grants access to all relevant information found in the Malmquist Productivity Index literature. Therefore if one wants to know the whole picture about the change in technology –benchmark (virtual) and best practice– and efficiency –comprising technical and scale efficiency–, choosing (19) would ease such task since all terms are directly calculated or can be easily determined by simple computations, *e.g.* productivity change at optimal scale $\mathrm{PTC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$, efficiency change $\mathrm{EC}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$ as well as returns to scale $\mathrm{RTS}_{\mathrm{O}}^{1,2}\left(x_i^1, y_i^1, x_i^2, y_i^2\right)$.

6. Empirical analysis

In this section we present the results reported by Simar and Wislon (1998:19) regarding productivity change in 17 OECD countries. Here we stress the role of the scale –bias– of technical change when interpreting technological and efficiency change. The database consists of annual figures on labor, capital and gross domestic product for 17 countries, which are taken from the Penn World Tables (version 5.6) and have been previously used by FGNZ (1994) and RD (1997). Table 2 shows the geometric mean of all indices that have been proposed in the literature over the 12 periods (1979-80, 1980-81,…, 1989,1990), while the geometric mean for all countries is reported in the last row.

There we identify an average annual productivity gain of 0.67%. On average we see that the leading countries, which are productively efficient from a technical and scale perspective, drive potential productivity gains to a 0,42% per year. Since average productivity change exceeds potential productivity change, we conclude that a catching up process in OECD countries exists, which is equivalent to an efficiency gain of 0.25% per year, *i.e.* $\mathrm{EC}_{\mathrm{O}}^{1,2}$ = $\breve{\mathrm{M}}_{\mathrm{O}}^{1,2} / \mathrm{PTC}_{\mathrm{O}}^{1,2}$ –eq. (7) by FGLR(1989,1994). This increase in efficiency is explained by a better productive performance both in technical and scale terms. In fact, we see that the average efficiency gain is mainly explained by a converging process toward optimal scale, as the average scale efficiency change index $\mathrm{SEC}_{\mathrm{O}}^{1,2}$ yields a 0.18% annual increase; two and a half times greater than technical efficiency change $\mathrm{TEC}_{\mathrm{O}}^{1,2}$ that reaches 0.07%, and showing how countries also get closer on average to the best practice frontier –eq. (17) by FGNZ (1994).

Given the importance of the scale efficiency change component when explaining the annual 0.25% productive efficiency gain, it is important to determine what its sources are. Since $\mathrm{SEC}_{\mathrm{O}}^{1,2} = \mathrm{RTS}_{\mathrm{O}}^{1,2} / \mathrm{STC}_{\mathrm{O}}^{1,2}$, the converging process toward optimal scale is sustained by the existence of increasing returns to scale –which contribute with an average 0.05% annual productivity gain, fostered by a change in the scale of the technology which on average works in favor of OECD countries –a 0.10% annual value –eq. (13). In fact, from the average country input scale perspective and the consecutively updated base periods, countries tend to get closer –by way of increasing returns to scale– to each period's optimal scale –mainly represented by the U.S., which is normally responsible for the shift in the benchmark virtual technology. But contemporarily these optima show a convergence toward the average country reference input scale –a favorable scale bias of technical change as discussed in section 3.1.1. The fact that changes in the scale of the technology works in favor of these countries input scales may be equivalently shown by comparing average potential productivity change at optimal scale $\mathrm{PTC}_{\mathrm{O}}^{1,2}$ to the average technical change value, $\mathrm{TC}_{\mathrm{O}}^{1,2}$. In this case, productivity gains coming from shifts in the best practice technologies at these countries' input scales beat those of the leading productively efficient countries by a 0.10% per year, $\mathrm{STC}_{\mathrm{O}}^{1,2} = \mathrm{PTC}_{\mathrm{O}}^{1,2} / \mathrm{TC}_{\mathrm{O}}^{1,2}$ –eq. (16). Therefore, the growth differential between these figures would reflect how technical change shows a scale bias favorable in average to OECD countries, supporting the scale convergence of 0.18% per year previously shown, and which is mainly responsible for the overall average efficiency gain.

It is now possible to turn our attention to the relevant sources responsible for the average productivity gain of 0.67% per year, which can be found in the rate of technical progress, $\mathrm{TC}_{\mathrm{O}}^{1,2}$, technical efficiency gain $\mathrm{TEC}_{\mathrm{O}}^{1,2}$ and the contribution of increasing returns to scale $\mathrm{RTS}_{\mathrm{O}}^{1,2}$ –eq. (18) by RD (1997). The shift in the best practice frontier shows an average technical progress of 0.55%. Nevertheless, there is a catching–up process of 0.07% per year since the average productivity gain represented by the original CCD (1982) Malmquist index exceeds technical progress by such amount –eq. (8). Finally, as previously discussed, the contribution of increasing returns to scale equals 0.05% annually.

Table 2. The Malmquist Productivity Index and its decompositions;
Geometric means (1979-1990)

	Product. Change	Potential Product. Change	Effi. Change	Tech. Change	Tech. Eff. Change	Returns to Scale	Scale Eff. Change	Scale of Tech. Change
Country	$\breve{M}_O^{1,2}$	$PTC_O^{1,2}$	$EC_O^{1,2}$	$TC_O^{1,2}$	$TEC_O^{1,2}$	$RTS_O^{1,2}$	$SEC_O^{1,2}$	$STC_O^{1,2}$
Australia	1.0112	1.0104	1.0007	1.0094	0.9990	1.0028	1.0017	1.0010
Austria	0.9947	0.9989	0.9957	1.0085	0.9995	0.9868	0.9962	0.9905
Belgium	1.0153	1.0105	1.0049	1.0096	1.0030	1.0027	1.0019	1.0008
Canada	1.0150	1.0105	1.0045	1.0104	1.0036	1.0010	1.0009	1.0001
Denmark	1.0033	0.9990	1.0044	1.0101	0.9976	0.9957	1.0068	0.9890
Finland	1.0257	1.0105	1.0150	1.0075	1.0107	1.0073	1.0043	1.0029
France	1.0115	1.0101	1.0013	1.0100	1.0011	1.0003	1.0002	1.0001
Germany	1.0072	1.0104	0.9968	1.0105	0.9966	1.0002	1.0002	0.9999
Greece	0.9982	0.9987	0.9996	0.9991	0.9981	1.0010	1.0015	0.9996
Ireland	1.0066	1.0000	1.0066	–	1.0000	–	1.0066	–
Italy	0.9995	0.9946	1.0049	0.9949	1.0048	0.9998	1.0001	0.9997
Japan	1.0010	0.9979	1.0031	0.9965	1.0003	1.0042	1.0028	1.0014
Norway	1.0154	1.0105	1.0049	1.0108	1.0000	1.0046	1.0049	0.9997
Spain	0.9967	0.9993	0.9973	0.9996	0.9970	1.0000	1.0003	0.9997
Sweden	1.0129	1.0104	1.0025	1.0073	1.0000	1.0056	1.0025	1.0031
UK	0.9982	0.9982	1.0000	0.9982	1.0000	1.0000	1.0000	1.0000
USA	1.0021	1.0021	1.0000	1.0058	1.0000	0.9963	1.0000	0.9963
ALL	1.0067	1.0042	1.0025	1.0055	1.0007	1.0005	1.0018	0.9990

Source: Own elaboration from Simar and Wilson (1998)

7. Conclusions

In the last decade several Malmquist indexes definitions and decompositions have been proposed in the literature. Each set of authors supported their own proposals criticizing the weaknesses of the opposing views, but never tried to find the common ground that would render all terms meaningful from a theory of production context.

We first show how each one of the different terms in which the Malmquist productivity index can be decomposed may be interpreted consistently, assigning alternative and non-competing names to each one of them, *e.g.* potential benchmark technical change $PTC_O^{1,2}$ versus actual best practice technical change $TC_O^{1,2}$, or scale efficiency change $SEC_O^{1,2}$ versus returns to scale, $RTS_O^{1,2}$. In doing so we overcome the concept shortage that limited the understanding and proposals of several authors, while giving room and valuing all terms that have been proposed in the literature. Also, we show how the competing decompositions of FGNZ (1994) and RD (1997) are linked by the concept of the scale –bias– of technical change. Introduced by Simar and Wilson (1998) and Zofio and Lovell (1998) in two widely known but so far unpublished working papers, this concept has a clear interpretation and enables us to show all relevant information with regard to a firm's own productivity change, as well as to that of the leading firms which, at the end, are responsible for technological change.

Finally, this leads us to conclude that a decomposition of the MPI that includes the scale bias of technical change term would enrich the analysis, allowing a complete assessment of the general framework where productivity change, as well as technological and efficiency change –both from a technical and a scale perspective–, take place. Hence we believe that the decomposition introduced by Simar and Wilson (1998) and Zofío and Lovell (1998) provides a unifying framework where one may deal with a complete characterization of technological and efficiency change. How such complete analysis can be undertaken is illustrated with a set of OECD countries.

Bibliography

Balk, B. (2001) Scale Efficiency and Productivity Change, *Journal of Productivity Analysis*, 15, 153-183.

Banker, R. (1984) Estimating Most Productive Scale Size Using Data Envelopment Analysis, *European Journal of Operations Research*, 17, pp. 35–44.

Caves, D., L. Christensen, and E. Diewert (1982a) The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity, *Econometrica*, 50(6), 1.393-1.414.

Färe, R. and S. Grosskopf (1998) Malmquist Productivity Indexes: A Survey of Theory and Practice, in *Essays in Honor of Sten Malmquist* (Eds.) R. Färe, S. Grosskofp and R. Russell. Dordrecht: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1989, 1994) Productivity Developments in Swedish Hospitals: A Malmquist Output Approach in *Data Envelopment Analysis: Theory, Methodology and Applications* (Eds.) A. Charnes, W. Cooper, A. Lewin and L. Seiford (1994), Dordrecht: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994) Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries, *American Economic Review*, 84 (1), 66-83.

Färe, R., S. Grosskopf, and M. Norris (1997) Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Reply, *American Economic Review*, 87 (5), 1040-1043.

Färe, R. y D. Primont (1995a), *Multi-output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Dordrecht.

Farrell, M. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*. Serie A, General, 120 (3), 282-284.

Førsund, F. (1997) The Malmquist Productivity Index, TFP and Scale, Working Paper, Department of Economics and Business Administration, University of Oslo.

Grosskopf, S. (2003) Some Remarks on Productivity, *Journal of Productivity Analysis*, 20, 459-474.

Grifell-Tatjé, E. and C.A.K. Lovell (1995) A Note On the Malmquist Productivity Index, *Economic Letters*, 47, 169-175.

Grifell-Tatjé, E. and C.A.K. Lovell (1996, 1999) A Generalized Malmquist Productivity Index, *Top*, 7 (1), 81–101.

Lovell, C.A.K. (2003) The Decomposition of Malmquist Productivity Indexes, *Journal of Productivity Analysis*, 20, 437-438.

Nishimizu, M. and J. Page (1982) Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, *Economic Journal*, 92, Dec., 920-936.

Orea, L. (2002) Parametric Decomposition of a Generalized Malmquist Productivity Index, *Journal of Productivity Analysis*, 18, 5-22.

Olesen, B. and N. Petersen (2003), Introduction, *Journal of Productivity Analysis*, 20, 237-241.

Ray, S. (1998) Measuring Scale Efficiency from a Translog Production Frontier, *Journal of Productivity Analysis*, 11, 167-182.

Ray, S. (2001) On an Extended Decomposition of the Malmquist Productivity Index, paper presented at the Seventh European Workshop on Efficiency and Productivity, Universidad de Oviedo, Sept. 25-29, Oviedo.

Ray. S. and E. Desli (1997) Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment, *American Economic Review*, 87 (5), 1.033-1.039.

Shephard, R. (1970), *Theory of Cost and production Functions*, New Jersey: Princeton University Press,

Simar, L. and P.W. Wilson (1998) Productivity Growth in Industrialized Countries, Discussion paper nº 9810, Universite Catholique de Louvain, Belgium.

Wheelock, D. and P. Wilson (1999), Technical Progress, Inefficiency, and Productivity Change in U.S. Banking, 1984-1993, *Journal of Money, Credit and Banking*, 31, 212-234.

Zofio, J.L. and C.A.K. Lovell (1998) Yet Another Malmquist Productivity index Decomposition, Mimeo, Department of Economics, Universidad Autónoma de Madrid, Madrid, University of Georgia, Athens, Georgia