ON THE CIRCULARITY OF THE MALMQUIST PRODUCTIVITY INDEX

by

Finn R. Førsund[*]

Department of Economics,

University of Oslo

and

The Frisch Centre, Oslo

email: f.r.forsund@econ.uio.no

June 2002

Abstract: Circularity is a desirable property of a productivity index seldom satisfied in available bilateral indices, such as the Malmquist index. Within a setting of micro units belonging to groups with group-specific frontier technology, the bilateral Malmquist productivity index is investigated. Our setting can be interpreted as representing cross section data, but also cross section, time series by identifying groups as time periods. A general proposition giving the condition for the Malmquist index to be circular is presented. When the condition is not met, ways of making the index transitive are explored. Four strategies adopted from the literature is followed; using one group as base, taking an average over possible bases, developing a multilateral index, and chain-linking. An expression of the Malmquist multilateral index is developed giving the difference between the Malmquist bilateral- and multilateral indexes.

Key Words: Circularity, transitivity, Malmquist productivity index, multilateral index, chain-linking.

JEL Classification: D24

[*] The paper has been written while a visiting fellow at ICER autumn / spring 2001/2002. I am indebted to Rolf Aaberge for valuable comments and encouragement.

1. Introduction

The Malmquist productivity index introduced in Caves, Christensen and Diewert (1982a) has grown in popularity the last decade (see Färe, Grosskopf and Russell (1998) for a review of applications and references to more than 100 papers on the index). We will here be concerned with investigating various aspects of the property of *circularity* of this bilateral index.

The construction of indexes for studying productivity is based on the general theory of indexes. The axiomatic approach to index number theory is to specify a number of properties an index should have, and then examine candidates for index formulas by applying the so- called *tests* to check if the desirable properties are fulfilled. According to Caves, Christensen and Diewert (1982b), p.74 "One of the principal issues in the index number literature early in this century was whether use of various indexes gave rise to transitive comparisons." Transitivity was regarded as one of the fundamental properties an index should obey. Indeed, it still should be in the field of productivity indexes. In general, if production unit *k* is more productive than unit *f*, and unit *f* is more productive than unit *l*, then unit *k* should be more productive than unit *l*. More specifically in our context transitivity allows for a *unique* ranking of units according to productivity. This may obviously be important for policy purposes at a micro level, and also of interest when comparing group aggregates such as countries. In a time series setting, covering events such as going from regulation to deregulation, one would be interested in comparing productivity development from different regimes.

Comparing productivity levels of units belonging to different groups, the *circular test* in the axiomatic index literature means that if we have an index for the comparison of productivity between units *k* and *f*, and between *l* and *f*, we can establish a productivity comparison between units *k* and *l* via the arbitrary third unit, *f*, that is independent of which third unit, f, that is chosen.

Notice that transitivity is not identical to circularity: circularity is sufficient but not necessary for transitivity. Reviewing the axiomatic test approach to index theory, Samuelson and Swamy (1974), p. 576, expressed the importance of the circular test as follows:

> Conclusion: So long as we stick to the economic theory of index numbers, the circular test is as required as is the property of transitivity itself.

From the axiomatic index literature we know that transitivity is impossible to combine with the other most desirable properties of an ideal index such as Fisher's[1]. *Characteristicity* (see Drechsler, 1973) has been used as a term to indicate the degree to which weights are specific to the comparison at hand. The Fisher ideal index utilises weights that are perfectly characteristic. As pointed out in Drechsler (1973) characteristicity and circularity are always in conflict with each other. Some degree of characteristicity must be sacrificed to obtain circularity[2].

In Diewert (1987), p.773, four strategies to follow if circularity does not hold are mentioned. Translated into our setting of units belonging to groups the strategies are:

i) choose one group as a base
ii) take an average over all possible choices as base
iii) abandon the use of a bilateral formula and develop an entirely new multilateral approach
iv) use the chain principle.

We will follow this programme and show how to adapt the Malmquist productivity index accordingly.

[1] Fisher did not easily accept this. As Samuelson and Swamy (1974, p.575) expressed it: "Indeed, so enamoured did Fisher become with his so-called Ideal index ..... that, when he discovered it failed the circularity test he had the hubris to declare "..., therefore, a *perfect* fulfilment of this so-called circular test should really be taken as proof that the formula which fulfils it is erroneous" … . Alas, Homer has nodded; or more accurately, a great scholar has been detoured on a trip whose purpose was obscure from the beginning." The Fisher view is still alive: realising that the Malmquist index they use is not transitive, Färe et al. (1994b) call Fisher's argument for the unimportance of circularity for "convincing" (p.80, footnote 22). It is, of course, legitimate to use an index that does not obey circularity, but the consequences should be investigated.

[2] "To be characteristic requires that each bilateral comparison ignore the "outside world". However, the "outside world" is always something else from bilateral comparison to bilateral comparison; and if one uses different weights, i.e. different yardsticks in each bilateral comparison, one cannot expect the requirement of circularity to be met." (Drechsler (1973), p.20).

In the literature an example of approach i) for the Malmquist index is found in Berg et al. (1993). In a setting similar to ours Nordic banks are studied by assuming separate technologies, and then, by using the frontier for one country as a common reference, productivity between countries are compared by comparing the efficiency scores of the largest banks in each country, as well as the average banks. A common Nordic technology was also tried. Using a fixed reference frontier technology yields a transitive index, but this index is dependent on the technology chosen (see also Berg, Førsund and Jansen (1992), Førsund (1990), (1993) for further comments on this index).

Following approach iii) for a constant returns to scale translog transformation function Caves, Christensen and Diewert (1982b) operated with unit-specific technologies (first-order parameters unit specific and second order parameters independent of unit), and developed a *multilateral* transitive productivity index for a comparison of two units, involving all other units. For each of the two units to be compared the unit's productivity relative to all other units is calculated as the geometric mean of the bilateral productivity comparison between the unit in question and each of the other units. These two comparisons are then combined to yield a multilateral transitive productivity index. The productivity comparison was based on proportional adjustment of outputs so that each country's outputs could be producable by the observed inputs using the other country's technology. A drawback pointed out is that the index has to be recomputed for new units.

The Caves, Christensen and Diewert (1982b) multilateral index is transitivised in such a way that a minimum of characteristicity is lost, conforming to the so called *EKS procedure* of minimising the difference between the transitive index and an ideal index (see Drechsler (1973), pp. 28-29, Caves, Christensen and Diewert (1982b), p. 83, Balk and Althin (1996), pp. 23-24). It is therefore of particular interest to develop a multilateral transitive Malmquist index following the Caves, Christensen and Diewert (1982b) approach. The last sentence of Caves, Christensen and Diewert (1982a) provides additional motivation:

> The index numbers that we have proposed in this paper, for arbitrary scale economies, can be extended to multilateral comparisons following the approach recommended in our previous paper. [Caves, Christensen and Diewert (1982b)]

To the best of our knowledge this has not been done satisfactorily in the literature yet for the Malmquist productivity index.

The closest attempt may be in Balk and Althin (1996). There another multilateral approach to obtain circularity is apparently advocated for cross section time series data. Instead of accommodating a two-period index concept to a multi-period setting, it is claimed that it is preferable to look at the measurement of productivity, efficiency and technical change in a multi-period setting from the outset. A multiplicative decomposition into an efficiency change term and a frontier change term is established directly without reference to the Malmquist productivity index, and the proposed index is not a Malmquist index. The solution to the transitivity problem of the frontier term is to calculate an index of the shift from one period to another as the geometric mean over distances between these two frontiers calculated for all observations in all time periods. A special feature is that the frontier shift term then is independent of the unit being compared for two time periods. Unit specificity is only present through the efficiency term. As was the case for the Caves, Christensen and Diewert (1982b) index this index has to be recomputed for new units or new time periods.

The plan of the paper is as follows: Section 2 reviews the definition of the Malmquist index and points to the reason for circularity not being fulfilled in general, and a theorem for when circularity holds is proved. It is shown that both simultaneous homotheticity and constant returns to scale, and Hicks neutrality and constant returns to scale are sufficient for circularity to hold. Four approached to transitivising the Malmquist productivity index is developed in Section 3, including fixed base, average base, multilateral and chain-linking. Further comments on the literature are offered in Section 4, and Section 5 concludes.

2. The Malmquist productivity index

The Malmquist productivity index, introduced in Caves, Christensen and Diewert (1982a), is a binary comparison of two entities, in empirical applications usually the same unit at different points in time, but we may also compare different units at the same point in time. We will formally use the latter, and operate with two different units. We will consider units belonging to subgroups having the same frontier production technology. Usually in productivity studies the specific subgroups represent different years for a (possibly unbalanced) panel of the units in question. In our general setting one limiting case is that the units are completely different between the subgroups. The opposite limiting case is that the group represents a time period, and that we have a panel, i.e. the units are the same for all groups. In a cross section context the group may be a geographical region like a country, or activities that have separate characteristics such as technology or ownership, that make them different, but still of interest and relevance to compare[3]. We may want to compare both productivity levels between groups in some average sense, and productivity at the micro level of the units (e.g. firms).

For a formal statement, consider a set of groups, T, with a total number of groups being $\#T$. Each group has a specific frontier technology. A group t (t Î T) consists of a set of N_t units with a total number of units, n_t . The total number of units across all groups is $\sum_{t \in T} n_t = n$. Each unit has a subscript for type of group (or technology), i.e. m_t is unit m belonging to group t, m_t Î N_t. The general production technology for a group, t, is expressed by the following production possibility set:

$$S_t = \{(y,x): x \text{ can produce } y\},\ t \in T \tag{1}$$

where y is the vector of outputs and x the vector of inputs.

The Farrell input-oriented technical efficiency measure, E_1, coincides with the inverse of the Shephard (1953) input distance function, and the output-oriented measure, E_2, coincides with the output distance function:

$$\begin{aligned} E_{1,t}(y,x) &= Min_q \{q : (y, qx) \in S_t\} \\ E_{2,t}(y,x) &= Min_j \left\{ j : (\frac{y}{j}, x) \in S_t \right\},\ t \in T \end{aligned} \tag{2}$$

[3] Homogenous outputs and inputs must be assumed in general, i.e. aluminium plants in different locations (countries) may be compared, but not aluminium plants and ice-cream factories.

Consider two units, k_u and l_v, where the subscripts u, v indicate the group membership (u,v Î T). The *Bilateral Malmquist productivity* index, $M_{d,u}(k_u,l_v)$, introduced in Caves, Christensen and Diewert (1982a) is, in our setting for units k_u and l_v with frontier technology for group u as base, defined by:

$$M_{d,u}(k_u,l_v) = \frac{E_{d,u}(l_v)}{E_{d,u}(k_u)} \ ,$$

$$\text{where } \ E_{d,t}(h_j) \equiv E_{d,t}(y^{h_j}, x^{h_j}), \tag{3}$$

$$d = 1{,}2 \ , \ k_u \in N_u \ , l_v \in N_v \ , h_j \in N_j \ , \ u,v,j,t \in T$$

To simplify the expressions we just indicate the unit in question to represent output- and input quantities within the Malmquist- and efficiency measure functions as shown in the last expression in (3). The type of orientation and technology are shown in the subscripts. As to which unit to be entered in the numerator and denominator on the rhs of the index we will follow the convention of having the efficiency evaluation of the unit entered first in the Malmquist index in the denominator and the second in the numerator. Thus unit l_v is more productive than unit k_u if $M_u(k_u,l_v) > 1$, and vice versa.

Productivity in the case of *input-orientation* (d =1) is defined by maximal reduction of inputs, i.e. a ratio of the minimal uniform input deflation factors such that the input-corrected observations lie on the production surface of one of the two technologies involved, here for group u. *Output-oriented* (d =2) productivity is defined by a ratio of maximal uniform output expansion factors such that the output-corrected observations lie on one of the two production surfaces, here technology u. The deflation and expansion factors correspond to the inverse of Shephard (1953) distance functions used by Caves, Christensen and Diewert (1982a). Thus, E_{1t} is the input deflation factor and E_{2t} is the inverse of the output expansion factor. By using Farrell efficiency measures the form of the expression for the Malmquist index is independent of type of orientation, d = 1,2. We will therefore drop the subscript for type of orientation in the following.

Circularity

Berg, Førsund and Jansen (1992) and Førsund (1990), (1993) pointed out that the Malmquist index (3) is not circular in general. In order to study the reasons for this lack of circularity we will first state the circular test in general terms.

DEFINITION 1: CIRCULARITY

Consider a bilateral index function, *I*, and the three values *I(k,l)*, *I(k,f)* and *I(l,f)* for the units *k, l, f.* The index satisfies the *circular test* if the following expression is valid:

$$I(k,l) = \frac{I(k,f)}{I(l,f)} \ \forall k,l,f \tag{4}$$

If we set $l = k$ in (1) we have that a natural normalisation called the *identity test* in the axiomatic index literature is also fulfilled; $I(k,k) = 1$. Further, it follows immediately that $I(k,l) = 1/ I(l,k)$. The requirement for circularity may then also be expressed as $I(k,l) = I(k,f)\, I(f,l)$.

It is straightforward to see that an index satisfying the circularity test must be transitive. If we adopt the convention that unit *l* is more productive than unit *k* if $I(k,l) > 1$(in accordance with the convention for (3)), and assume that $I(k,f) < 1$, meaning that unit *k* is more productive than unit *f*, and $I(f,l) <1$, meaning unit *f* is more productive than unit *l*, then it follows from (4) that $I(k,l) < 1$, meaning unit *k* is more productive than unit *l*, i.e. transitivity is preserved.

The Malmquist index satisfies the identity test. Setting $l_v = k_u$ in the definition (3) we have that the numerator and denominator on the rhs become equal. Further, we have that $M_u(k_u, l_v) = 1/M_u(l_v, k_u)$ by interchanging the units in the definition (3).

The general requirement for a Malmquist productivity index to be circular, is set out in the following theorem:

THEOREM 1

Consider the bilateral Malmquist index defined in (3) for units k_u and l_v, and the Farrell efficiency measures defined in (2). Introducing an arbitrary third unit, f_w, the Malmquist index is circular according to definition (4) if and only if

$$\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)} = 1,\ k_u \in N_u, l_v \in N_v, f_w \in N_w, w,u,v \in T \tag{5}$$

Proof:

Applying the definition (4) of circularity to the Malmquist index (3) and substituting the Farrell efficiency measures and rearranging terms we get:

$$M_u(k_u,l_v) = \frac{M_u(k_u,f_w)}{M_v(l_v,f_w)} = \frac{E_u(f_w)/E_u(k_u)}{E_v(f_w)/E_v(l_v)} = \frac{E_v(l_v)}{E_u(k_u)}\frac{E_u(f_w)}{E_v(f_w)} =$$

$$\frac{E_u(l_v)}{E_u(k_u)}\left[\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)}\right] = M_u(k_u,l_v)\left[\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)}\right], \tag{6}$$

$$k_u \in N_u, l_v \in N_v, f_w \in N_w, w,u,v \in T$$

The last expression in (6) is the lhs of (5). Equation (6) can only be valid if (5) is true, thus the “if” part is established.

To see that condition (5) implies transitivity, start with the definition (4) of the Malmquist index between units k_u and l_v in terms of efficiency terms. Multiplying the expression $E_u(l_v)/E_u(k_u)$ with (5) leaves it unchanged, since the expression in (5) has the value 1:

$$M_u(k_u,l_v) = \frac{E_u(l_v)}{E_u(k_u)} = \frac{E_u(l_v)}{E_u(k_u)}\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)} = \frac{E_u(f_w)/E_u(k_u)}{E_v(f_w)/E_v(l_v)} = \tag{7}$$

$$\frac{M_u(k_u,f_w)}{M_v(l_v,f_w)},\ k_u \in N_u, l_v \in N_v, f_w \in N_w, w,u,v \in T$$

The last equality establishes the circularity according to definition (4).

The expression on the lhs of (5) can be interpreted as composed of relative distances between isoquants of the two technologies *u* and *v*. In the numerator the ratio of efficiency scores for observation f_w measured against technologies u and v expresses the relative distance between isoquants for the input-or output levels of the observation. The same interpretation holds for the denominator, but now the relative distance is measured along the ray through the observation l_v and a corresponding change of inputs- or outputs. We see that the crucial factor for circularity is that the two expressions for relative distance between the frontiers for group u and group v at the two observations f_w and l_v cancel out. In general, the Malmquist index will not be transitive[4]. We can rephrase Theorem 1 as requiring that relative distances between isoquants of the two technologies involved in the bilateral comparison must be the same for all possible third observations.

The requirement for circularity may be illustrated further by following Färe et al. (1994a) and splitting the Malmquist bilateral productivity index into an *efficiency change* term (or catching-up term), EC_{uv}, and a *frontier change* term, FC_{uv}[5]:

$$M_u(k_u,l_v) = \frac{E_u(l_v)}{E_u(k_u)} = \frac{E_v(l_v)}{E_u(k_u)}\frac{E_u(l_v)}{E_v(l_v)} = EC_{uv}(k_u,l_v)FC_{vu}(l_v),$$

where (8)

$$EC_{uv}(k_u,l_v) = \frac{E_v(l_v)}{E_u(k_u)}\ ,\ FC_{vu}(l_v) = \frac{E_u(l_v)}{E_v(l_v)}$$

$$k_u \in N_u, l_v \in N_v, u,v \in T$$

The efficiency term, EC_{uv}, is circular. This may be established in the same way as above by introducing a third unit, f_w, from group w. Then using condition (4) and the definition of the efficiency change term in (8) we get:

$$E_{uv}(k_u,l_v) = \frac{E_{uw}(k_u,f_w)}{E_{vw}(l_v,f_w)} = \frac{E_w(f_w)/E_u(k_u)}{E_w(f_w)/E_v(l_v)} = \frac{E_v(l_v)}{E_u(k_u)}, \tag{9}$$

$$k_u \in N_u, l_v \in N_v, f_w \in N_w, w,u,v \in T$$

which shows that the efficiency change term is circular. It is the frontier change term,

[4] The statement in Färe et al. (1994b) that (3) is transitive is obviously a misunderstanding, as pointed out in Balk and Althin (1996), footnote 6.

[5] Restating the decomposition in Nishimizu and Page (1982) done for parametric frontier functions for non-parametric ones.

FC_{uv}, that is not circular, and then neither the Malmquist index itself. The root of the non-circularity problem is what is revealed by the decomposition in (8): the frontier change term is based on the relative distance between isoquants from the two technologies involved measured at the same observation (l_v in (8)). Using the definition (4) of circularity to the frontier change term defined in (8) we get:

$$FC_{vu}(l_v) = \frac{FC_{wu}(f_w)}{FC_{wv}(f_w)} = \frac{E_u(f_w)/E_w(f_w)}{E_v(f_w)/E_w(f_w)} = \frac{E_u(f_w)}{E_v(f_w)}, \tag{10}$$

$$l_v \in N_v,\ f_w \in N_w, w,u,v \in T$$

The last expression is not the correct definition of the frontier change term: the arbitrary observation f_w has taken the place of the observation l_v involved in the bilateral comparison. Circularity then requires that:

$$\frac{E_u(f_w)}{E_v(f_w)} = \frac{E_u(l_v)}{E_v(l_v)} \Rightarrow \frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)} = 1, l_v \in N_v,\ f_w \in N_w, w,u,v \in T \tag{11}$$

After rearranging terms in the first equation we arrive at exactly the same requirement as stated in Theorem 1.

An illustration

The general situation is illustrated in Figure 1 for two outputs. Three observations, k_u, l_v and f_w , and (six) *factor isoquants* (a Frisch (1965) concept), corresponding to the input levels used by observations k_u , l_v and f_w and belonging to the two technologies, u and v, are shown. We need two isoquants in general corresponding to the two technologies for each observation. The observations k_u , l_v (and f_w) are inefficient. The catching-up component is:

$$EC_{vu}(k_u, l_v) = \frac{E_v(l_v)}{E_u(k_u)} = \frac{Ol_v / Ob_v}{Ok_u / Oa_u} \tag{12}$$

The efficiency of observation l_v measured against the frontier isoquant of technology v corresponding to the observed resources, x^{l_v}, is Ol_v/Ob_v and the efficiency of observation k_u measured against the frontier isoquant of technology u corresponding to

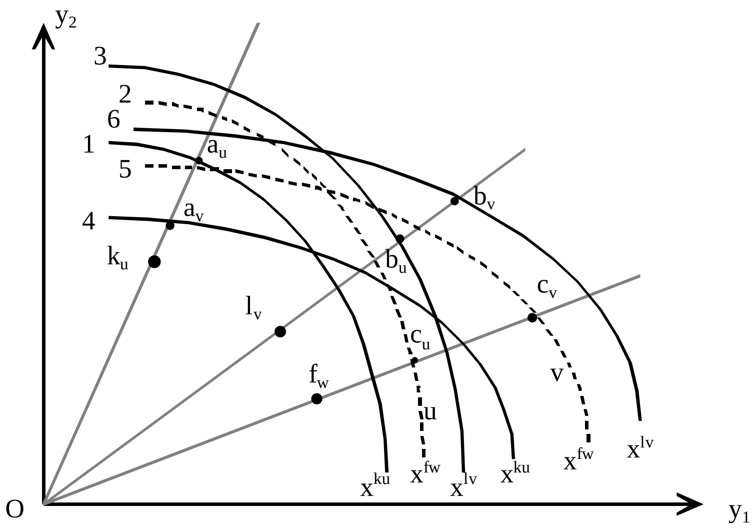

Figure 1. Distances between input isoquants

the observed resources, x^{k_u} , is Ok_u/Oa_u. In order to check circularity a third observation, f_w, is also indicated in the figure) and the two isoquants (dotted curves from *u* and *v* technologies corresponding to the resources, x^{f_w}.

The frontier shift term is the relative distance between the relevant factor isoquants measured for the output ratio of observation l_v :

$$F_{uv}(l_v) = \frac{E_u(l_v)}{E_v(l_v)} = \frac{Ol_v / Ob_u}{Ol_v / Ob_v} = \frac{Ob_v}{Ob_u} \tag{13}$$

The last expression illustrates the property of the frontier shift measure of being based on the relative distance between frontiers. The isoquants in question for the two technologies both correspond to the resource level, x^{l_v}, of observation l_v . As to the problem with circularity we see that the relative distance between isoquants from the two technologies corresponding to the resource level, x^{f_w}, of the third (arbitrary) observation is Oc_v/Oc_u , which in general is different from the relative distance in (13). We have from (11) that equality is required for circularity.

The circularity deviation index

The expression in (5) involving the relative distances between the two frontiers may be developed into an indicator, *D*, for relative deviation from circularity. The third observation, f_w, is arbitrary, so we have to run through all possible third observations. Since ratios are involved, taking a geometric mean is appropriate:

$$D(k_u, l_v) = 1 - \frac{\left[\prod_{w \in T} \left[\prod_{f_w \in N_w} E_u(f_w)/E_v(f_w) \right]^{1/n_w} \right]^{1/\#T}}{E_u(l_v)/E_v(l_v)} =$$

$$1 - \frac{\left[\prod_{f \in N} E_u(f)/E_v(f) \right]^{1/n}}{E_u(l_v)/E_v(l_v)} \quad , \; k_u \in N_u, l_v \in N_v \,, f_w \in N_w \,, f \in N, \; w,u,v \in T \tag{14}$$

To derive the last expression we have simplified by letting *f* ($f \in N$) represent a general unit. Comparing the expression with (8) we see that the last term corresponds to the geometric mean of the frontier change terms. For perfect circularity $D = 0$, while a positive value indicates that the geometric mean of the relative distance between the isoquants for the frontiers *u* and *v* for output ratios of all the observations is smaller than the relative distance for the two isoquants involved for unit l_v for its output ratios, and vice versa for a negative value of *D*. We see that in Figure 1 the contribution from observation f_w alone on the *D*-value will be negative, since $Oc_v/Oc_u > Ob_v/Ob_u$. By taking geometric means over all possible pairs of k_u and l_v we can also develop a *global* indicator for deviation from circularity. The exercise is left to the reader.

The geometric mean-based Malmquist index

The basic definition (3) of the Malmquist index may seem more symmetrical if we involve the technology of both observations in the bilateral comparison. The geometric mean of the two indices with the two technologies as base is proposed in Färe et al. (1994a) and there termed the Malmquist productivity index[6] :

$$\overline{M}_{uv}(k_u, l_v) = \left[M_u(k_u, l_v) M_v(k_u, l_v) \right]^{1/2}, \; k_u \in N_u \,, l_v \in N_v \,, u,v \in T \tag{15}$$

[6] This somewhat unfortunate practice, because it may be confused with the original proposal, is followed by many in the literature. (Färe et al. (1994a) circulated widely as a working paper from 1989).

Concerning the decomposition (8) it is straightforward to see that the efficiency term, EC_{uv}, remains the same, and that the frontier change term $FC_{vu}(l_v)$ now is the geometric mean of the two relative distances between the technologies measured at each of the two observations, $[FC_{vu}(l_v)\, FC_{uv}(k_u)]^{1/2}$:

$$\begin{gathered}
\overline{M}_{uv}(k_u, l_v) = EC_{vu}(k_u, l_v)\overline{FC}_{uv}(l_v, k_u) , \\
EC_{uv}(k_u, l_v) = \frac{E_v(l_v)}{E_u(k_u)} , \ \overline{FC}_{vu}(l_v, k_u) = \left[\frac{E_u(l_v)}{E_v(l_v)} \frac{E_v(k_u)}{E_u(k_u)}\right]^{1/2} , \\
k_u \in N_u , l_v \in N_v , \ u,v \in T
\end{gathered} \tag{16}$$

However, a problem with this version is that a mean over just two components may hide interesting structural variation, while being far from a mean in the sense of appealing to the law of large numbers for stability. In Figure 1 the u technology is more efficient than the v technology at output ratios of observation l_v, and vice versa at output ratio of observation k_u. This crucial information gets lost taking the average.

As regards circularity the geometric mean version has the same problem as the basic definition (3). The problem that the relative distance between the relevant isoquants for the technologies u and v involved in the bilateral comparison for a third observation must be equal to the distance between the isoquants for one of the observations of the bilateral comparison does not go away. In fact, we now also get a requirement on the distance between isoquants for the w technology and the u- and v technologies respectively. Using the circularity definition (4) on the geometric mean Malmquist index (16), using the definitions in (8), the requirement in (5) now reads:

$$\begin{gathered}
\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)} \cdot \frac{E_w(k_u)/E_v(k_u)}{E_w(l_v)/E_v(l_v)} = 1 , \\
k_u \in N_u , l_v \in N_v , \ f_w \in N_w , w,u,v \in T
\end{gathered} \tag{17}$$

The first term appears in (5), and the second term is due to using the geometric mean version of the Malmquist index. It is now not so straightforward to establish Theorem 1. It cannot be excluded at the outset that there are combinations of relationships between isoquant distances that result in the value of 1 for the total expression. However, (17) is not demanding more than (5). If we use the Malmquist index definition (3) and inspect the bilateral index between f_w and l_v, using k_u as the third arbitrary observation, then the requirement (5) in Theorem 1 is just that the second expression in (17) must be 1.

The homothetic case

Is it possible that the underlying production possibilities (1) may be structured in such ways that the Malmquist index (3) is circular? A general smooth multi-output multi-input production function, *F(y,x)*, with standard neo-classical properties may be used to describe the efficient border of the production set in (1). Provided some standard regularity condition are fulfilled, we may according to McFadden (1978) write (1) in an equivalent way:

$$S_t = \{(y,x) : x \ can \ \ produce \ \ y\} \equiv \{(y,x) : F_t(y,x) \le 0\}, \ t \in T \tag{18}$$

Here $F(y,x) = 0$ represents the efficient border of the set S_t .

We will show that if the frontier production functions in (18) all are *simultaneous homothetic* as defined by Hanoch (1970) and exhibit *constant returns to scale* then the circularity test (4) is fulfilled.

DEFINITION 2: INPUT- AND OUTPUT HOMOTHETICITY

Following Hanoch (1970) input- and output homotheticity is defined by the existence of functions $y(m,y)$ and $f(m,x)$ with $y(1,y) = 1$ and $y_m' > 0$, and with $f(1,x) = 1$ and $f_m' > 0$, such that:

$$\begin{aligned} &\text{Input homotheticity}: F[my, y(m,y)x] = 0 \ , \\ &\text{Output homotheticity}: F[f(m,x)y, mx] = 0 \\ &\text{for } m > 0 \text{ and } (y,x) \text{ satisfying } F(y,x) = 0 \end{aligned} \tag{19}$$

The set of isoquant surfaces $\{x^* : F(my_o, x^*) = 0\}$ is derived from the isoquant surface defined by $\{x : F(y_o, x) = 0\}$, by a uniform expansion of each input $x_i^* = y(m,y)x_i$ for a uniform increase, m, of all outputs y_o. Correspondingly, the set of transformation surfaces $\{y^* : F(y^*, mx_o) = 0\}$ is derived by uniform expansion of each output $y_j^* = f(m,x)y_j$ for a uniform increase, m, in the inputs.

The scale elasticity function is special for homothetic functions. The elasticity of scale, e, can be defined in general either by considering a proportional change in inputs and then calculating the elasticity of the proportional change in outputs, or considering a proportional change in outputs and then calculate the inverse elasticity of the proportional change in inputs:

$$e(y,x) = \frac{\partial b(m,y,x)}{\partial m}\frac{m}{b} = \frac{1}{\frac{\partial a(m,y,x)}{\partial m}\frac{m}{a}} \tag{20}$$

Carrying out the differentiation after inserting $b(m,y,x) = f(m,x)$ for output homotheticity and $a(m,y,x) = y(m,y)$ for input homotheticity, evaluating, without loss of generality, the derivatives at $m = 1$, we get:

$$\frac{\partial b(1,y,x)}{\partial m} \equiv e(y,x) = -\frac{\sum_{n=1}^{N}\frac{\partial F(y,x)}{\partial x_n}x_n}{\sum_{m=1}^{M}\frac{\partial F(y,x)}{\partial y_m}y_m} = \frac{\partial f(1,x)}{\partial m} = e^x(x) \tag{21}$$

$$\frac{1}{\frac{\partial a(1,y,x)}{\partial m}} \equiv e(y,x) = -\frac{\sum_{n=1}^{N}\frac{\partial F(y,x)}{\partial x_n}x_n}{\sum_{m=1}^{M}\frac{\partial F(y,x)}{\partial y_m}y_m} = \frac{1}{\frac{\partial y(1,y)}{\partial m}} = e^y(y) \tag{22}$$

DEFINITION 3: SIMULTANEOUS HOMOTHETICITY

Simultaneous homotheticity is defined when both input- and output homotheticity according to Definition 2 holds at the same time. When input- and output homotheticity is fulfilled at the same time, we must have:

$$f_m'(1,x) = \frac{1}{y_m'(1,y)} \Rightarrow e^x(x) = e^y(y) \tag{23}$$

In the input homothetic case the contour curves of the scale elasticity function coincides with output isoquants, and in the output homothetic case the contour curves coincide with input isoquants.

We need a definition of what is meant by the frontier technology for each group belonging to the same family of homothetic functions:

DEFINITION 4: FAMILY OF HOMOTHETIC FUNCTIONS

Consider a homothetic function $F(y,x) = 0$ satisfying (19) in Definition 2. We will then define functions $F_u(y,x)$ ($u \in T$) as belonging to the same family if:

$$\begin{aligned}&\left\{x_o^* : F_u(y_u, y_{uo}(1, y_u)x_u) = 0\right\}\\&\left\{y_o^* : F_u(f_{uo}(1, x_u)y_u, x_u) = 0\right\}\\&\forall u \in T, o \in T, y_{uo}(1, y_u) = f_{uo}(1, x_u) = 1 \; for \; u = o\end{aligned} \tag{24}$$

where $y_o = y_u$ in the first set, and $x_o = x_u$ in the second set. The output- and factor isoquants are all radial projections of each other. Different technologies just mean different labelling of output-and input isoquants.

The expansion functions correspond to the efficiency measure functions, cf. the definitions (2), linking in general an inefficient point to the corresponding isoquant of the frontier function. In the case of two outputs the situation is portrayed in Figure 2. The input isoquants representing the two technologies, *u* and *v*, are radial projections of each other by definition. We can therefore calculate the output oriented efficiency measure either against the isoquant corresponding to the *u* technology or the *v* technology. The difference must be a factor of proportionality depending on the two technologies. Let us illustrate by using the observation f_w in Figure 1. The efficient isoquants are isoquant No. 2 for the *u* technology and No. 5 for the *v* technology. The efficiency measures against each of the isoquants are:

$$E_u(f_w) = \frac{Of_w}{Oc_u}, \; E_v(f_w) = \frac{Of_w}{Oc_v} \tag{25}$$

We can then write the efficiency measure relative to the *u* technology as:

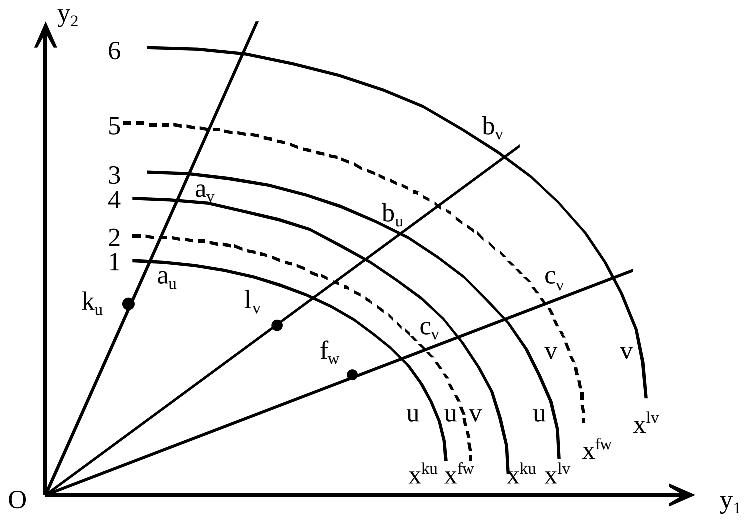

Figure 2. The homothetic case

$$E_u(f_w) = E_v(f_w)\frac{Oc_v}{Oc_u} \tag{26}$$

The efficiency score for an observation relative to a technology can be written as the product of the efficiency score relative to another technology and a correction factor depending on the two technologies. This must be true in general. The interesting question is under which conditions the correction factor is independent of the observation at hand. An answer is found in the following proposition:

PROPOSITION 1: EFFICIENCY MEASURES AND SIMULTANEOUS HOMOTHETICITY

Consider a frontier production function satisfying the conditions for simultaneous homotheticity according to the definitions (19) and (24). Assume that all group frontier production functions defined in (18) belong to this same class of homothetic functions, then the following holds:

i) Choosing a frontier technology, *o*, for group *o* as base the input- and output oriented Farrell efficiency measures defined by (2) can be written:

$$\begin{aligned} E_{1,u}(k_u) &= E_{1,o}(k_u)B_o(u, y^{k_u}) \\ E_{2,u}(k_u) &= E_{2,o}(k_u)A_o(u, x^{k_u}) \;\;, k_u \in N_u,\, u, o \in T \end{aligned} \tag{27}$$

ii) If the technology in addition exhibits constant returns to scale, then the input-oriented measure is independent of the output level of unit k_u, and the output-oriented measure is independent of the input-levels for unit k_u.

$$E_{1,u}(k_u) = E_{1,o}(k_u)B_o(u)$$
$$E_{2,u}(k_u) = E_{2,o}(k_u)A_o(u) \quad , k_u \in N_u, u,o \in T \tag{28}$$

iii) The mark-up factors for the input- and output oriented efficiency measures are identical:

$$B_o(u) = A_o(u) \equiv H_o(u) , k_u \in N_u, u,o \in T \tag{29}$$

Proof:

Part i): Consider the outputs and inputs observed for unit k_u symbolised by (y^{k_u}, x^{k_u}). The efficiency corrected output- and input levels considering output-orientation and input-orientation separately are, using definition (2), $((1/\varphi_u)y^{k_u}, \theta_u x^{k_u})$. We will choose technology o ($o \in T$) as a base technology. The set of inputs, x_o, belonging to the factor isoquant of the o technology for output level y^{ku} is:

$$\{x_o : F_o(y^{k_u}, x_o) = 0\} \tag{30}$$

Define the Farrell input efficiency measure, θ_o, for the observation k_u against frontier technology o by:

$$E_{1,o}(k_u) = Min_{\theta_o}\{\theta_o : F_o(y^{k_u}, \theta_o x^{k_u}) = 0\} \tag{31}$$

According to the definition (24) of frontier functions belonging to the same family of homothetic functions, we can express the set of inputs, x_o, also in terms of frontier technology, u:

$$\{x_o : F_u(y^{k_u}, \gamma_{uo}(1, y^{k_u})\theta_u x^{k_u}) = 0\} \tag{32}$$

Comparing (24), (25) and (26) we then have that:

$$x_o = \theta_o x^{k_u} = \gamma_{uo}(1, y^{k_u})\theta_u x^{k_u} \Rightarrow \theta_u = \theta_o \gamma_{uo}(1, y^{k_u})^{-1}$$
$$\Rightarrow E_{1,u}(k_u) = E_{1,o}(k_u)B_o(u, y^{k_u}) \tag{33}$$

The input-oriented efficiency score for a unit from a specific group may be obtained form the efficiency score relative to a base technology adjusted multiplicatively by a factor depending on the current and base technologies, and the output level of the observation. If the base technology is more efficient than the current frontier technology, then the adjustment factor is greater than one.

Defining output-oriented efficiency for observation k_u relative to the *o* technology we get:

$$E_{2,o}(k_u) = Min_{j_o}\left\{ j_o : F_o\left(\frac{y^{k_u}}{j_o}, x^{k_u}\right) = 0 \right\} \tag{34}$$

The output set for the levels $y_o = y^{ku}/j_o$ expressed by *u* technology is:

$$\left\{ y_o : F_u\left(f_{uo}(1, x^{k_u})\frac{y^{k_u}}{j_u}, x^{k_u}\right) = 0 \right\} \tag{35}$$

Comparing (34) and (35) we have that:

$$\begin{aligned} &\frac{y^{k_u}}{j_o} = f_{uo}(1, x^{k_u})\frac{y^{k_u}}{j_u} \Rightarrow j_u = j_o f_{uo}(1, x^{k_u}) \\ &\Rightarrow E_{2,u}(k_u) = E_{2,o}(k_u) A_o(u, x^{k_u}) \end{aligned} \tag{36}$$

Part ii) of the proposition is seen by the fact that constant returns to scale means that the function $F(y, x)$ is homogeneous of degree 1: $F(sy, sx) = sF(y, x)$, implying that $y(m,y) = m$, and $f(m,x) = m$. The relative changes in outputs and inputs must be the same.

Part iii) of the proposition follows straightforwardly from the fact (already mentioned in Farrell, 1957) that the oriented efficiency scores are identical under constant returns to scale.

COROLLARY 1: CIRCULARITY AND HOMOTHETICITY

> Assume that Proposition 1, part i), ii) and iii), is satisfied. Then the Malmquist productivity index satisfies the circular test (4).

Proof:

Inserting (28) and (29) from Proposition 1 in the requirement (4) in Theorem 1 for circularity yields :

$$\frac{E_u(f_w)/E_v(f_w)}{E_u(l_v)/E_v(l_v)} = \frac{E_o(f_w)H_o(u)/E_o(f_w)H_o(v)}{E_o(l_v)H_o(u)/E_o(l_v)H_o(v)} = 1\,, \tag{37}$$

$$k_u \in N_u, l_v \in N_v, f_w \in N_w, w,u,v,o \in T$$

REMARK 1:

It does not matter for circularity which base, *o*, we chose. E_o and H_o adjust keeping the product constant when the base, *o*, is changed.

We see from Figure 2 that when all the homothetic frontier functions belong to the same family, then the relative distance between the isoquants is independent of the output mix. In order for the spacing of isoquants not to have an impact on the relative distance the homothetic family has to be constant returns to scale.

DEFINITION 5: HICKS-NEUTRALITY AND CIRCULARITY

Consider production possibilities defined by (1) and assume constant returns to scale. The efficiency measures as defined in (2) for two different production possibility sets of the two technologies *u* and *o* (*u*,*o* Î *T*) are:

$$E_{1,u}(k_u) = E_{2,u}(k_u),\ \ E_{1,o}(k_u) = E_{2,o}(k_u)\ ,\ k_u \in N_u,\ u,o \in T \tag{38}$$

Then Hicks-neutrality for a comparison of two different production possibility sets is defined when the following holds for the efficiency measures of the two technologies *u* and *o*:(*u*,*o* Î *T*) (see Chambers and Färe (1994) for the standard time series case):

$$E_{1,u}(k_u) = E_{1,o}(k_u)H_o(u)\ ,\ E_{2,u}(k_u) = E_{2,o}(k_u)H_o(u)\ ,\ k_u \in N_u,\ u,o \in T \tag{39}$$

COROLLARY 2:

Homothetic functions satisfy Hicks neutrality.

Proof:

This is established directly by comparing (28) - (29) and (39).

REMARK 2:

Hicks neutrality has been defined for time series observations in the literature, but we can adopt it to our setting quite straightforwardly. There is a close connection between homotheticity and Hicks neutrality. It is tempting to conjecture that simultaneous homotheticity and constant returns to scale is equivalent with Hicks neutrality and constant returns to scale. But without further investigations we cannot exclude that other production functions may also yield efficiency measures that satisfy (28) - (29).

3.Transitivising the Malmquist productivity index

Restricting the underlying production function to be simultaneous homothetic, or exhibit Hicks neutrality, and constant returns to scale may be seen as too restrictive for many applications. We will therefore investigate the options suggested by Diewert (1987) for trasitivising the Malmquist index .

The fixed technology Malmquist index

Inspecting the decomposition in (5) there is an immediate way of making the Malmquist index transitive: allow only one technology to be used when defining the Malmquist indices for the observations k_u and f_w , and l_v and f_w . Then the relative distance term in the last expression in the second line of (5) becomes 1 by definition. If we impose technology *u* as fixed, then the efficiencies for the third observation, f_w, are now measured against the same technology as for the observation, l_v against technology *u*. But, of course, the circularity hinges on keeping the same technology.

The fixed technology need not be one of the two corresponding to the observations to be compared, but may be representing a third group. This will complicate the decomposition into an efficiency change term and a frontier change term. Introducing a common technology index, *o*, assumed to belong to the technology set, *T*, Berg et al. (1992) did the following decomposition into a catching-up term, E_{uv}, and a frontier-shift term, FC_{vuo}:

$$M_{Fo}(k_u, l_v) = \frac{E_o(l_v)}{E_o(k_u)} = \frac{E_v(l_v)}{E_u(k_u)} \frac{E_u(k_u)/E_o(k_u)}{E_v(l_v)/E_o(l_v)} =$$

$$EC_{uv}(k_u, l_v) FC_{vuo}(k_u, l_v) \quad (40)$$

where

$$EC_{vu}(k_u, l_v) = \frac{E_v(l_v)}{E_u(k_u)} ,\; FC_{uvo}(k_u, l_v) = \frac{E_u(k_u)/E_o(k_o)}{E_v(l_v)/E_o(l_v)}$$

$$k_u \in N_u, l_v \in N_v, u, v, o \in T$$

The Malmquist index has been given an extra subscript, *F*, in front of the technology subscript to distinguish the fixed technology index from the basic one, (3). The efficiency term, EC_{uv} , is identical to the same term in the Malmquist index (8), while there is a "double" relativity introduced in the frontier shift term. The distances between the frontier for the *u* and *v* technologies are measured relative to the *o* frontier for the observations k_u and l_v . It is this “double relativity” that leads to terms canceling out when we involve a third observation, f_w , checking for transitivity of *FC* following the procedure in (10). From earlier we have that the efficiency term is transitive, and from the last expression in (40) it may easily be established that also the new frontier change term is transitive. But again, the payment for transitivity is the dependency on the reference frontier technology, *o*.

In Diewert et al. (1982) it is also proposed to use a representative unit as a base. This can be interpreted as the same procedure as above. Representative just means that there are some reasons for picking a specific reference base. Using the analogy from standard fixed weigths- changing weights indexes (Laspeyre and Paasche) in a time series context using the first or the last year makes sense. In a cross section context the purpose of the study may point to a specific base, or the size of the group with the representative technology, or the superiority of the technology.

Since no averaging is performed to obtain transitivity, the fixed technology form is especially suitable if the interest is on following individual units appearing in several groups, as may be the case in cross section, time series data. The dependency on the fixed technology is less a drawback the more obvious the choice of a representative technology.

Using the average as a base

Diewert (1987) proposed to use an average over all possible bases. One way of averaging is to average the data and construct the technology for the average unit for the total sample. In a Malmquist index context this may not be so straightforward. Averaging of technologies does not seem so attractive either. But one way of establishing a base would be to pool all the data and establish a technology for this pooled set. Then using this set as a base will technically look like the formulation above in (40) with the pooled technology having the index, *o*. This procedure is similar to the notion of *inter temporal* technology in a time series context (Tulkens and van den Eechaut, 1995). For time series their notion of *accumulating* technology may also be used as an averaging procedure.

The multilateral Malmquist index

Caves, Christensen and Diewert (1982) developed a bilateral productivity index based on information from all units and technologies, and termed this index a multilateral index. They start out by using the geometric mean version of the bilateral index, stating that they find this "natural", because it is then base-invariant. But as we have commented above, it is only base invariant as to the two technologies involved, and taking the mean may distort or conceal interesting information. We will therefore not follow this practice of taking geometric mean at this stage. It will be shown that the key development in Caves, Christensen and Diewert (1982b) does not depend on taking this mean.

Following the approach in Caves, Christensen and Diewert (1982b), developed for the translog transformation function, for the Malmquist productivity index (3) we will compare the productivity of unit k_u and unit l_v , respectively, with all other units m_t , $t \in T$. Then going through all possible bilateral comparisons we get the two geometric mean indices, $\overline{M}_u , \overline{M}_v$, for units k_u and l_v respectively. For notational simplification we will also use the notation m for a unit in general suppressing the technology index, t, and then letting this index run over all possible units in the set *N*. The index for the relative productivity of unit k_u and unit l_v is based on the geometric means of all possible bilateral comparisons for both units:

$$\overline{M}_u(k_u) = \left[\prod_{t\in T}\left(\prod_{m_t\in N_t} M_u(k_u,m_t)\right)^{1/n_t}\right]^{1/\#T} = \left[\prod_{m\in N} M_u(k_u,m)\right]^{1/n},\ k_u \in N_u, u\in T \quad (41)$$

$$\overline{M}_v(l_v) = \left[\prod_{t\in T}\left(\prod_{m_t\in N_t} M_v(l_v,m_t)\right)^{1/n_t}\right]^{1/\#T} = \left[\prod_{m\in N} M_v(l_v,m)\right]^{1/n},\ l_v \in N_v, v\in T \quad (42)$$

We first take the geometric mean for all the bilateral Malmquist indexes between the units k_u and l_v and the unit m_t , and then run through all the groups, *#T*, thus ending up with taking the geometric mean over all possible bilateral comparisons.

The multilateral firm Malmquist productivity index, $MT(k_u,l_v)$, can now be formed:

$$MT(k_u,l_v) = \frac{\overline{M}_u(k_u)}{\overline{M}_v(l_v)} = \frac{\left[\prod_{m\in N} M_u(k_u,m)\right]^{1/n}}{\left[\prod_{m\in N} M_v(l_v,m)\right]^{1/n}} = \left[\prod_{m\in N} M_u(k_u,m)M_v(m,l_v)\right]^{1/n} \quad (43)$$

$k_u \in N_u, l_v \in N_v, u,v\in T, m\in N$

The index (43) is transitive. Consider a new unit f_w belonging to the group with w -technology. The transitive multilateral indexes between k_u and f_w , and l_v and f_w are, using the basic definition:

$$MT(k_u, f_w) = \frac{\overline{M}(k_u)}{\overline{M}(f_w)},\ MT(l_v, f_w) = \frac{\overline{M}(l_v)}{\overline{M}(f_w)} \quad (44)$$

The general requirement of transitivity (4) is:

$$MT(k_u,l_v) = \frac{MT(k_u, f_w)}{MT(l_v, f_w)} \quad (45)$$

Inserting the first equality of (43) on the right-hand side of (44) we have:

$$\frac{MT(k_u, f_w)}{MT(l_v, f_w)} = \frac{\overline{M}(k_u)/\overline{M}(f_w)}{\overline{M}(l_v)/\overline{M}(f_w)} = MT(k_u,l_v) \quad (46)$$

The general formula can also be decomposed into an efficiency part and a frontier change part. Inserting the definitions of the efficiency- and frontier change terms in (8), we have after some straightforward manipulations:

$$MT(k_u,l_v)=\left[\prod_{m\in N}\frac{E_u(m)}{E_u(k_u)}\frac{E_v(l_v)}{E_v(m)}\right]^{1/n}=\frac{E_v(l_v)}{E_u(k_u)}\left[\prod_{m\in N}\frac{E_u(m)}{E_v(m)}\right]^{1/n}= \tag{47}$$

$$EC_{vu}(k_u,l_v)\overline{FC}_{uv}(m)\ ,\ k_u\in N_u,l_v\in N_v,u,v\in T,m\in N$$

The efficiency term remains the same as in the general definition (8), while the frontier change term is the geometric mean for the relative distance between the isoquants for u- and v-technology measured over all observations m (including k_u and l_v).

One way of assessing the change implied by transitivising the Malmquist index can be obtained by taking out from the geometric mean expression in (43) the term we get when setting $m = l_v$, i.e. letting the two units coincide:

$$MT(k_u,l_v)=EC_{uv}(k_u,l_v)\left[\frac{E_u(l_v)}{E_v(l_v)}\prod_{m\in N,m\neq l_v}\frac{E_u(m)}{E_v(m)}\right]^{1/n}=$$

$$EC_{uv}(k_u,l_v)FC_{vu}(k_u,l_v)\left[\prod_{m\in N,m\neq l_v}\frac{E_u(m)}{E_v(m)}\right]^{1/n-1} \tag{48}$$

The last term measures the change of the original Malmquist index (3) due to transitivising the bilateral firm productivity index by introducing the multilateral term involving the comparisons of the two units k_u and l_v with all the other units except unit l_v. Note that (48) corresponds to the last term in the circularity deviation index (14).

We have seen that the Caves, Christensen and Diewert 1982b) idea of transitivising by forming geometric means of all possible bilateral indexes for two units can be done without starting with a geometric version (15). However, it may be of interest to also show the bilateral geometric mean versions, as will become evident when we comment upon the literature in Section 4. A notational inconvenience is that since in this case all the technologies will be explicitly involved, we have to use the first expressions in (41) and (42):

$$\overline{M}_g(k_u) = \left[\prod_{t \in T}\left(\prod_{m_t \in N_t}\left[M_u(k_u, m_t)M_t(m_t, k_u)\right]^{1/2}\right)^{1/n_t}\right]^{1/\#T} =$$

$$\left[\prod_{t \in T}\left(\prod_{m_t \in N_t}\overline{M}_{ut}(k_u, m_t)\right)^{1/n_t}\right]^{1/\#T}, \; k_u \in N_u, m_t \in N_t, u, t \in T \tag{49}$$

where the last expression results after inserting (15). The geometric mean based index, $\overline{M}_g(l_v)$, for unit l_v is quite similar to (49), and is not shown. The multilateral index can now be written:

$$\overline{MT}(k_u, l_v) = \frac{\overline{M}_g(k_u)}{\overline{M}_g(l_v)} = \frac{\left[\prod_{t \in T}\left[\prod_{m_t \in N_t}\overline{M}_{ut}(k_u, m_t)\right]^{1/n_t}\right]^{1/\#T}}{\left[\prod_{t \in T}\left[\prod_{m_t \in N_t}\overline{M}_{vt}(l_v, m_t)\right]^{1/n_t}\right]^{1/\#T}} =$$

$$\left[\prod_{t \in T}\left[\prod_{m_t \in N_t}\overline{M}_{ut}(k_u, m_t)\overline{M}_{tv}(m_t, l_v)\right]^{1/n_t}\right]^{1/\#T}, \; k_u \in N_u, l_v \in N_v, u, v \in T, m \in N \tag{50}$$

By repeating the procedure for establishing circularity applied to (43) we have that this version of the multilateral index is also satisfying circularity.

We can also do the decomposition (47). The efficiency term remains the same, while the frontier change term will contain the the bilateral geometric means defined in (16):

$$\overline{MT}(k_u, l_v) = EC_{uv}(k_u, l_v)\left[\prod_{t \in T}\left[\prod_{m_t \in N_t}\overline{FC}_{tu}(k_u, m_t)\overline{FC}_{vt}(m_t, l_v)\right]^{1/n_t}\right]^{1/\#T} \tag{51}$$

It is straightforward also to reproduce Equation (48) to show the realtion between the bilateral geometric Malmquist index (16) and the corresponding multilateral one, (50).

Now, what is the difference between the formulations (43) and (50)? A first observation is that while the efficiency change terms in (48) and (51) have maximal charactericity, depending only on the two units k_u and l_v, the frontier shift term in (48) has *minimal* charactericity, since the general unit, *m*, only once takes the values of k_u, respectively l_v. But looking at the bilateral geometric mean variant we have from (15) or (16) that the units k_u and l_v are participating in defining distances between all the isoquants for the technologies *u* and *t* and *v* and *t* respectively for $t \in T$. This ”over-representation” (compared with the index (43)) of the units k_u and l_v when calculating distances only concerns the frontier change terms in (51). As mentioned in Section 1 the EKS procedure may be used for deriving the most preferred transitivised index. To check whether this is form (43) or (50) is outside the ambition of this paper[7].

Chain-linking

There is a way of obtaining transitivity by *chain-linking* (Caves, Christensen and Diewert, 1982b). The idea of chain-linking is to utilise an ordering of the units to build up the bilateral index between two units multiplicatively by using the bilateral indexes for the units within the ordering between the two units under consideration in a succesive fashion. The adjacent observations are compared directly, while non-adjacent observations are compared only indirectly, using the intervening observations as intermediaries. Let us introduce an ordering between the units. We assume that units k_u and l_v are not adjacent. The chain-linked index, M_{CH} (k_u , l_v) between the units k_u and l_v ordered such that k_u comes before l_v can then be written:

$$M_{CH}(k_u, l_v) = M_u(k_u, a_a) M_a(a_a, b_b) ......M_z(z_z, l_v),\ \ u, v, a, b, .., z \in T \qquad (52)$$

We have assumed that unit k_u is adjacent to a_a, and then a_a to b_b, etc., until unit z_z which is the unit before unit l_v in the general ordering. We may apply the basic definition (3) for the intermediate expressions between adjacent units in (52).

Let us now introduce a unit f_w that is further out in the ordering than unit l_v. To check if the chain index (52) is circular we inspect the following expression:

[7] Drechsler (1973) formulates the EKS procedure as taking a weighted geometric mean of the bilateral characteristic index in question (double weight) and any possible combination of two chain indirect indices (single weight). Balk and Althin (1996) set up the formal minimising expression.

$$M_{CH}(k_u,l_v)=\frac{M_{CH}(k_u,f_w)}{M_{CH}(l_v,f_w)}=\frac{M_u(k_u,a_a)M_a(a_a,b_b)......M_g(g_g,f_w)}{M_v(l_v,m_m)M_m(m_m,n_n)......M_g(g_g,f_w)}= \quad (53)$$

$$M_{CH}(k_u,l_v)\ , u,v,a,b,..,m,n,g,w\in T$$

Unit g_g is next before unit f_w, and unit m_m is the one after l_v, and then comes n_n, etc. When we cancel the same elements in the numerator and denominator we are just left with the elements in the chain from k_u to l_v. Changing the ordering between k_u, l_v and f_w yields the same result; the chain index (52) is circular. We just use that in general we have $M_u(k_u, l_v) = 1/M_u(l_v, k_u)$. By using this relation that follows from the identity test being fulfilled, the units can be entered such that the unit from the first ranking is entered first in the superscript on the Malmquist index (i.e. the unit first in the ranking is always in the numerator of (3)).

The circularity depends on the existence of a complete ordering, and on keeping this fixed. In the case of time series data we can interpret the groups, *t*, as representing time periods (e.g. years). In this case there is a natural ordering of the groups, but not of the units within groups. What is most common to do is to consider a unit being present in all time periods. A productivity comparison for this unit observed at two different time periods can be build up by chain-linking the adajent productivity indexes in the way descibed in (52).

However, note that the chain version for panel data is not transitive in a more general sense involving another unit than *k*, e.g. *l*. The expression $M_u(k_u,l_v)$ is, simply, not defined since it is only the technologies that are ordered.

4. Comments on the literature

The Caves, Christensen and Diewert papers (1982a) and (1982b) are closely linked. In the former paper a unit's productivity relative to another is defined as the maximal proportional increase in the outputs of the second unit such that the resulting output vector is producable with the second unit's input levels and the technology of the first unit. In the latter paper the bilateral productivity index in Eq. (2) there is a comparison of the output-oriented efficiency of the second unit with respect to the first unit's technology relative to the output-oriented efficiency of the first unit with respect to its own technology, where efficiency is defined as the maximal proportional increase in the output vector for given input vectors. We see that these definitions coincide when each unit is efficient with respect to its own technology, as assumed in Caves, Christensen and Diewert (1982a). In the case of efficient units in the sense just mentioned it is no surprise that the geometric mean of the Malmquist index over the two possible choices of technology, i.e. equation (15), is shown to be the Törnquist index when scale is corrected for. Caves, Christensen and Diewert (1982b) establish that the geometric mean of the two definitions of productivity obtained by changing the base unit is equal to the translog bilateral productivity index, assuming constant returns to scale and translog technologies with unit-specific first order terms, but unit independent second order terms, so that the quadratic identity of Diewert (1976) can be employed. The same assumptions and procedure is followed in Caves, Christensen and Diewert (1982a).

The approach of Balk and Althin (1996) is to establish a multilateral index directly by formulating an efficiency change term and a frontier shift term in a setting of cross section, time series data and a frontier production function for each period. However, in doing so the connection to the Malmquist index (3) is not mentioned. Let us reformulate their frontier shift term using efficiency measures instead of input distance functions, and use the subscript t for the time periods:

$$PR(k_t, k_{t'}) = EC(k_t, k_{t'})TC(k_t, k_{t'}) = \frac{E^t(k_t)E^{t'}(x)}{E^{t'}(k_{t'})E^t(x)}\ ,\ k_t \in N_t, k_{t'} \in N_{t'}, x \in N \qquad (54)$$

where x is an observation of an arbitrary unit. Now, letting $x = k_{t'}$ we have:

$$PR(k_t, k_{t'}) = \frac{E^t(k_t)E^{t'}(k_{t'})}{E^{t'}(k_{t'})E^t(k_{t'})} = \frac{E^t(k_t)}{E^t(k_{t'})},\ k_t \in N_t, k_{t'} \in N_{t'} \qquad (55)$$

Comparing this expression with (3) we see that this is the expression for the Malmquist productivity index measuring the shift between the frontiers for periods t and t' in the observation k_t. Inserting $x = k_t$ in (54) yields the Malmquist index (3) with the shift measured in observation $k_{t'}$. The Balk and Althin productivity index, PR, has a connection to the Malmquist index of the form (3) without making it independent of the base periods by taking the geometric mean.

Balk and Althin (1996) set out to define a new multilateral productivity index not being extensions of existing bilateral ones. It is somewhat puzzling to observe that they do not realise the connection to the bilateral Malmquist index. In fact, their proposal (54) is identical to our Equation (43) reinterpreted for a panel in a time series context. Furthermore, they develop our version (53) with reference to Caves, Christensen and Diewert (1982b), and even report calculations with their own index and (53) without pointing out that the difference is due to their version (54) not being based on first taking a geometric mean of the bilateral Malmquist index as in (15)[8].

5. Conclusions

Transitivity is an obvious requirement for index calculations when unique rankings are required. In a most general setting of units belonging to groups with common frontier technology, it has been demonstrated that the general requirements for the Malmquist productivity index to be circular are quite limiting as to choice of form of the production function. The four approaches to transitivise an index have been followed up. They all work, but have quite different characteristics. The fixed base Malmquist index is the simplest, also from a calculation point of view, since the basic definition may be used with just a slight modification. However, the weakness of this procedure is that the index depends on the technology chosen as base. This made Balk and Althin claim that "it appears to be not a productivity index at all" (p. 26). In light of the comments in Section 4, the judgement is left to the reader.

An alternative proposal in Caves, Christensen and Diewert (1982b) is to measure productivity relative to a representative unit. In our setting we take this suggestion as a confirmation that the approach of a fixed technology base for a bilateral Malmquist

[8] It is also odd that they report that the Caves, Christensen and Diewert (1982b) index variant they derive is, according to the EKS method, optimal. Why then stick to their own index? However, it should be pointed out that EKS optimality is not claimed in Caves, Christensen and Diewert (1982b).

index is valid. Our setting can be interpreted as cross section data, but also cross section, time series by identifying groups as time periods. We suggest that the fixed technology index may be preferable if one is interested in following single units in a cross section, time series setting.

The main idea of Caves, Christensen and Diewert (1982b) is to measure productivity for each of the two units involved in the bilateral comparison relative to all units in all groups and form the geometric means, thus turning the bilateral index into a multilateral one. Some characteristicity has to be sacrificed. Expressions for the Malmquist multilateral index have been developed giving the difference between the Malmquist bilateral and multilateral indexes in the cases of starting with the basic bilateral definition, and starting with the geometric mean for the two technologies involved, as favoured by Rolf Färe and associates. To chose between them the EKS procedure may be employed.

According to Caves, Christensen and Diewert (1982b) the multilateral index is attractive for cross section comparisons and for panel data, but not necessarily preferable to chain-linked bilateral indexes for time series data. As have been demonstrated the chain-linked bilateral index is transitive given that a complete ordering of all units makes sense. In the common panel data time series setting only time, i.e. the group technology, is ordered. This implies that chaining only for the same unit is circular.

The amount of calculations involved may also be considered in the choice of an index. The index based on a base technology or a representative unit is obviously the easiest to calculate. One strength of the Malmquist productivity index is that it allows following individual units. The use of all observations in order to obtain transitivity may lead to too much averaging out of individual developments in a cross section, time series setting.

References

Balk, B. M. and R. Althin (1996): ”A new, transitive productivity index”, *Journal of Productivity Analysis* 7, 19-27.

Berg, S. A., F. R. Førsund and E. S. Jansen (1992): "Malmquist indices of productivity growth during the deregulation of Norwegian banking", *Scandinavian Journal of Economics* 94, Supplement, 211-228.

Berg, S. A., F. R. Førsund, L. Hjalmarsson and M. Suominen (1993): "Banking efficiency in the Nordic countries", *Journal of Banking and Finance*, 17, 371-388.

Caves, D.W., L.R. Christensen and E. Diewert (1982a): "The economic theory of index numbers and the measurement of input, output, and productivity", *Econometrica* 50, 1393-1414.

Caves, D. W., L. R. Christensen and W. E. Diewert (1982b): "Multilateral comparisons of output, input, and productivity using superlative index numbers”, *Economic Journal* 92 (March), 73-86.

Chambers, R. G. and R. Färe (1994): “Hicks neutrality and trade biased growth: a taxonomy”, *Journal of Economic Theory* 64, 554-567.

Diewert, W. E. (1976): "Exact and superlative index numbers ", *Journal of Econometrics* 4, 115-145.

Diewert, W. E. (1987):”index numbers”, in Eatwell, J., M. Milgate and P. Newman (eds.): *The new Palgrave. A dictionary of economics*, London and Basingstoke: The Macmillan Press Limited, 767-780.

Drechsler, L. (1973): ”Weighting of index numbers in mulitlateral international comparisons”, *Review of Income and Wealth* 19, 17-34.

Frisch, R. (1965): *Theory of production*, Dordrecht: D. Reidel.

Färe, R., S. Grosskopf and R.R. Russell (eds) (1998): *Index numbers: Essays in Honor of Sten Malmquist*, Boston/London/Dordrecht: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1994a): “Productivity developments in Swedish hospitals: a Malmquist output index approach”, in Charnes, A., W. W. Cooper, A.Y. Lewin and L. M. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology, and Applications*, Boston/Dordrecht/London: Kluwer Academic Publishers, 253-272.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994b): “Productivity growth, technical progress and efficiency change in industrialized countries”, *American Economic Review* 84(1), 66-83.

Førsund, F. R. (1990): “The Malmquist productivity index”, *Memorandum No 28* from Department of Economics University of Oslo.

Førsund, F. R. (1993): "Productivity growth in Norwegian ferries", in Fried, H., C. A. K. Lovell, and S. Schmidt (eds.): *The measurement of productive efficiency: techniques and applications*, Oxford University press, 352-373.

Hanoch, G. (1970): "Homotheticity in joint production", *Journal of Economic Theory* 2, 423-426.

McFadden, D. (1978): "Cost, revenue and profit functions", in M. Fuss and D. McFadden (eds.): *Production economics: A dual approach to theory and applications*, Vol. 1, Chapter 1, 3-109, Amsterdam: North Holland Publishing Company.

Nishimizu, M. and J. M. Page (1982): “Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia 1965 – 78”, *Economic Journal* 92, 920-936.

Samuelson, P. A. and S. Swamy (1974): "Invariant economic index numbers and canonical duality: survey and synthesis", *American Economic Review* 64, 566-593.

Shephard, R. W. (1953): *Theory of cost and production functions*, Princeton: Princeton University Press.

Tulkens, H. and P. van den Eeckaut (1995): "Non-parametric efficiency, progress, and regress measures for panel data: methodological aspects", *European Journal of Operational Research* 80, 474-499.