Problems with (Dis)Aggregating Productivity, and Another Productivity Paradox

by

Kevin J. Fox*

The University of New South Wales
School of Economics
Sydney, NSW 2052, Australia.
E-mail: K.Fox@unsw.edu.au
Fax: +61-2-9313-6337
Tel: +61-2-9385-3320

ROUGH DRAFT
February 2002

Abstract

Using a standard definition of productivity growth, a country may have higher productivity growth than another country in each sector, but may have a lower productivity growth rate overall. This observation has significant implications for the aggregation and disaggregation of productivity growth estimates, and the interpretation of productivity convergence studies that have used cross-country sectorial data. In addition, it is shown that an increasingly popular method for aggregating sectorial estimates of productivity growth has a serious problem—it fails a basic test from index-number theory. This leads to problems for the interpretation of previously published estimates of e.g., contributions to aggregate productivity change from changes in industry structure. An index-number method that avoids these aggregation problems is introduced.

Key words: Productivity measurement, index numbers.

JEL classification: C43, D24, E23

---

* The author gratefully acknowledges the hospitality of the Universitat Autònoma de Barcelona.

1 Introduction

Aggregation problems have long bothered economists in various contexts. An area which has perhaps received relatively little attention has been the area of the aggregation of productivity. The main reference in this area is still Domar (1961). This paper demonstrates an aggregation paradox and examines the two currently most common methods for aggregating productivity, before introducing two alternative methods.

The potential problems in aggregation are illustrated by a "productivity paradox," which is that even using the same definition of productivity growth at different levels of aggregation can lead to paradoxical results. For example, one country may have higher productivity growth than another in every sector, yet have lower productivity growth overall. The aggregation method used in this case has been used repeatedly in empirical studies for international productivity comparisons, so this is an observation of some importance. The reason this result arises is illustrated, and it is noted that the result contains information about the reasons for the relative aggregate performance of countries.

The two aggregation methods examined in section 3 are used in the literature as a starting point for decompositions of aggregate productivity growth into contributions from changes in industry structure. Unfortunately, it is found here that both of the aggregation methods have the serious problem of not satisfying monotonicity. That is, for the same level of inputs, output can be higher for every firm in period t than period $t-1$ yet aggregate productivity growth can fall. This does not seem a very promising place to start when considering contributions from various sources to aggregate productivity change.

An alternative aggregation method is introduced which avoids the monotonicity problem of the other methods, and which has a particularly nice aggregation/decomposition properties. This method is based on the multiplicative Törnqvist (1936) index. Unfortunately, the method appears to fail when it is used to examine firm entry and exit issues. That is, it is very useful at the level of sectors or industries (where the same sector/industry exists in each period), but cannot adequately deal with examining a more detailed level of data which examines firm entry and exit contributions to productivity.

Hence, another alternative is introduced. This method is based on the additive Bennet (1920) indicator (or "index") (Diewert, 1998). It overcomes the problems inherent with the other methods, and provides an interesting decomposition into sources of productivity growth.

The paper is organised as follows. The next section demonstrates an aggregation paradox related to the construction of an aggregate productivity growth measure. Section 3 examines the two main aggregation methods used in examining contributions from changes in industry structure to aggregated productivity growth. Section 4 introduces a method which overcomes a problem inherent in the usually employed methods, but it is of limited use in examining detailed industry-level changes in structure. Section 5 introduces a method which appears to have very nice properties in terms of both aggregation and decomposition at various levels of interest. Section 6 concludes.

2 A Productivity Paradox

Consider the case, without loss of generality, where there are two countries A and B, and sectors 1 and 2 in each country. These sectors produce the same goods in each country. Now assume, for simplicity, that input growth same in both sectors in both countries. Hence, using a standard definition of total-factor productivity as output growth divided by input growth, output growth determines productivity growth.

Let Y_{ij}^t denote real value added in country i, $i = A, B$, sector j, $j = 1, 2$, for period t, $t = 0, 1$. Then consider the following case, where output growth (and hence productivity growth) is higher in each sector in country A than in country B. That is,

$$\frac{Y_{A1}^1}{Y_{A1}^0} > \frac{Y_{B1}^1}{Y_{B1}^0}, \tag{1}$$

and

$$\frac{Y_{A2}^1}{Y_{A2}^0} > \frac{Y_{B2}^1}{Y_{B2}^0}. \tag{2}$$

Then we can note the following paradox.

PARADOX *Although A has higher productivity growth in both sectors, it can have lower aggregate productivity growth than B. That is, the following is possible:*

$$\frac{Y_{A1}^1 + Y_{A2}^1}{Y_{A1}^0 + Y_{A2}^0} < \frac{Y_{B1}^1 + Y_{B2}^1}{Y_{B1}^0 + Y_{B2}^0}. \tag{3}$$

The aggregation of output across sectors by addition before using division to get growth rates should indicate an obvious potential for a troublesome result like this arising. However, the definition of productivity in (3) is consistent with the definition of productivity at the sectorial level. A simple re-expression of the first half of (3) suggests how this paradoxical result is possible, as follows:

$$\frac{Y_{A1}^1 + Y_{A2}^1}{Y_{A1}^0 + Y_{A2}^0} = \theta_{A1}^0 \cdot \frac{Y_{A1}^1}{Y_{A1}^0} + \theta_{A2}^0 \cdot \frac{Y_{A2}^1}{Y_{A2}^0}, \tag{4}$$

where θ_{Aj}^0 is the share of industry j in total output for country A, or $\theta_{Aj}^0 = Y_{Aj}^0/(Y_{A1}^0 + Y_{A2}^0)$, $j = 1, 2$. Naturally, a similar expression to (4) exists for country B. It is clear then that the sector shares play a role in determining aggregate productivity, and that they also play a role therefore in determining relative productivity between countries A and B.

Let g_{ij} denote productivity growth between periods 0 and 1, country i, sector j. Then using (4), equation (3) becomes:

$$\theta_{A1}^0 g_{A1} + \theta_{A2}^0 g_{A2} < \theta_{B1}^0 g_{B1} + \theta_{B2}^0 g_{B2}, \tag{5}$$

or

$$\theta_{A1}^0 g_{A1} - \theta_{B1}^0 g_{B1} < \theta_{B2}^0 g_{B2} - \theta_{A2}^0 g_{A2}. \tag{6}$$

Hence, if country A has relatively more of it's total output in the sector with lower growth, and country B has relatively more of it's total output in the sector with higher growth, then the paradoxical result is possible. Another simplifying assumption, and a numerical example illustrate this, as follows.

Let the output (and productivity) growth in sector 1 be the same in both countries, and similarly for sector 2; $g_{A1} = g_{B1} = g_1$ and $g_{A2} = g_{B2} = g_2$. Then (6) becomes

$$(\theta^0_{A1} - \theta^0_{B1})g_1 < (\theta^0_{B2} - \theta^0_{A2})g_2. \tag{7}$$

If $\theta^0_{A1} = 0.9$, $\theta^0_{A2} = 0.1$, $\theta^0_{B1} = 0.8$ and $\theta^0_{B2} = 0.2$:

$$(0.1)g_1 < (0.1)g_2 \Rightarrow 1 < g_2/g_1. \tag{8}$$

In the example above, productivity growth is higher in sector 2 than in sector 1, while productivity growth of each sector is the same in both countries, yet aggregate productivity growth is different because of the different shares. From this example, it is straightforward to see that the assumption of equal sectorial growth rates can be relaxed while the inequality in (6) still holds. That is, the inequality in the paradox holds because of A having a higher share of its economy than B in the sector with lower productivity growth.

What if input growth is not held constant across sectors and countries? It is unclear how to aggregate in this case. A common approach is to use output shares to weight productivity growth rates (as we will see in following sections). However, this is not a unique solution, as we will have an equation like (4) for inputs as well, and input shares could similarly be used as weights. Hence the choice of output or input shares is somewhat arbitrary.

The following sections examine issues relating to aggregating productivity over firms and industries within a country. It can be easily shown that all the methods considered below can produce paradoxical results (such as above) when used for cross-country comparisons. This is not necessarily a flaw with these methods, as they indicate performance in terms of allocation of resources (Fox, 1999). However, to avoid paradoxical results multilateral comparison techniques could be used to compare countries using cross-section or panel data; see e.g. Caves, Christensen and Diewert (1982a) and Diewert (1999).

3 Some Aggregation Methods

This section assesses two aggregation methods that have been proposed in the literature (Baily, Hulten and Campbell, 1992; Foster Haltiwanger and Krizan, 1998; Haltiwanger, 2000; Hahn, 2000). After aggregating productivity, the aggregate productivity measures are decomposed into contributing components. In the case of firm-level data, methods for obtaining contributions to aggregate productivity from changing industry structure and firm entry and exit have been proposed. Unfortunately, both methods for aggregating productivity have the serious problem of not satisfying the basic property of monotonicity, and hence the interpretation of the decompositions should be quite different to those currently given in the literature.

Let Y_n^t be an output aggregate for firm n, $n = 1, \cdots, N$, for periods $t = 0, 1$, let X_n^t be an input aggregate, and let θ_n^t be the share of firm n in total output, or some other appropriate weight; see e.g. Bartelsman and Gray (1996). Then define $\triangle TFP_G^{0,1}$ as the the ratio of the share-weighted geometric mean of firm total-factor productivity levels in period 1 relative to period 0:

$$\begin{aligned}
\triangle TFP_G^{0,1} &= \exp\left[\sum_{n=1}^{N}\theta_n^1 \ln TFP_n^1 - \sum_n \theta_n^0 \ln TFP_n^0\right] \\
&= \exp\left[\sum \theta_n^1 \ln(Y_n^1/X_n^1) - \sum \theta_n^0 \ln(Y_n^0/X_n^0)\right] \\
&= \exp\left[\left(\sum \theta_n^1 \ln Y_n^1 - \sum \theta_n^0 \ln Y_n^0\right) - \left(\sum \theta_n^1 \ln X_n^1 - \theta_n^0 \ln X_n^0\right)\right] \\
&= \prod\left[(Y_n^1)^{\theta_n^1}/(Y_n^0)^{\theta_n^0}\right] / \left[\prod (X_n^1)^{\theta_n^1}/(X_n^0)^{\theta_n^0}\right] \qquad (9)
\end{aligned}$$

The change in the total-factor productivity has been expressed as an index of output growth divided by an index of input growth in the last line of (9). Consider the output-growth index, $\prod[(Y_n^1)^{\theta_n^1}/(Y_n^0)^{\theta_n^0}]$. The problem with this as an index of output growth is that the weights on each output (the shares) are not held constant in going between periods 0 and 1, and hence it confounds quantity changes with share movements.

Another way of stating this is that it fails the monotonicity test from index-number theory (Diewert, 1993; p. 75). This test says that if the quantities increase between the two periods then the index should also increase. However, the quantity indexes in (9) clearly do not satisfy this property, as the changing shares between the two periods could result in $\sum \theta_n^1 \ln Y_n^1 < \sum \theta_n^0 \ln Y_n^0$ even though $Y_n^1 > Y_n^0$ for all j.[1]

An analogy can be drawn with the usual process by which a quantity index is constructed. With multiple goods, prices are often used as weights. Let p_m^t be the price of good m, $m = 1, \cdots, M$, in period $t = 0, 1$, and similar let q_m^t be the corresponding quantity vector. A quantity index, $Q^{0,1}$ can be written as follows:

$$Q^{0,1} = \frac{\sum_m p_m^t q_m^1}{\sum_m p_m^t q_m^0} \tag{10}$$

If p_m^t is set at the base-period level, p_m^0, then $Q^{0,1}$ is the well-known Laspeyres quantity index. If p_m^t is set at the current-period level, p_m^1, then $Q^{0,1}$ is the well-known Paasche quantity index. Taking the geometric mean of these two quantity indexes solves the problem of choosing between base-period and current-period prices, and yields Fisher's Ideal quantity index. If p_m^0 is used in the denominator and p_m^1 in the numerator, then $Q^{0,1}$ is a value-change index, *not* a quantity index. There are cases where one may be interested in such a value-change index, and its decomposition. Such a case is if the ratio of values represented the change in profits or nominal Gross Domestic Product between two periods. However, it would not be correct to call this a quantity index by virtue of its being the change in the value of quantity between the periods. The implied output and input "quantity indexes" in (9) have the same property as such a value-change index, in the sense that the weights on the quantities are allowed to change between the denominator and the numerator. The resulting indexes are not quantity indexes. Hence, equation (9) may still be of interest, but it is not an index of aggregate productivity growth.[2]

[1] Also, it can be shown invariance to units of measurement of this aggregation method only holds if shares in each sector are constant over time, or if weights sum to one. That is, the use of Domar weights (Domar, 1961), which do not sum to one, would make this method sensitive to the units of measurement.

[2] See Diewert (1992) for a derivation of a Fisher's Ideal productivity growth index from economic theory. See Caves, Christensen and Diewert (1982b) for the corresponding derivation for the Törnqvist index.

Another proposed method is very similar to the "geometric-mean" method of equation (9). In fact, authors typically cross-reference the decompositions arising from the above and following methods without even noting an important difference. That is, the following aggregation method uses TFP in levels rather than logarithms, and thus takes the difference in the share-weighted arithmetic means between the two periods:

$$\triangle TFP_A^{0,1} = \sum_{n=1}^{N} \theta_n^1 TFP_n^1 - \sum_{n=1}^{N} \theta_n^0 TFP_n^0. \tag{11}$$

Immediately a problem can be noted, as follows. Re-scale all outputs by λ:

$$\begin{aligned} \triangle \widetilde{TFP}_A^{0,1} &= \sum \theta_n^1 (\lambda Y_n^1 / X_n^1) - \sum \theta_n^0 (\lambda Y_n^0 / X_n^0) \\ &= \lambda \left[\sum \theta_n^1 (Y_n^1 / X_n^1) - \sum \theta_n^0 (Y_n^0 / X_n^0) \right] \\ &= \lambda \triangle TFP_A^{0,1}. \end{aligned} \tag{12}$$

That is, TFP can be increased by, e.g. changing the units of measurement of output. Hence, although the literature, and reviews of the literature (e.g. Balk, 2001), do not usually explicitly note this in mathematical form, empirical studies divide through by $\sum \theta_n^0 TFP_n^0$ in order to avoid this problem (e.g., Baily, Bartelsman and Haltiwanger, 2001; p. 424) . Thus, they get growth rates of TFP ($\triangle \widehat{TFP}$), which are then decomposed into components of interest. That is, using (11),

$$\triangle \widehat{TFP}_A^{0,1} = \frac{\sum \theta_n^1 TFP_n^1}{\sum \theta_n^0 TFP_n^0} - 1. \tag{13}$$

We noted above that the geometric-mean method of equation (9) had the problem with not satisfying the basic property of monotonicity. Now consider a re-expression of (13), assuming that inputs are constant, and dropping the (irrelevant for current purposes) subtraction of one:

$$\triangle \widehat{TFP}_A^{0,1} = \sum \theta_n^1 (Y_n^1 / X) / \sum_n \theta_n^0 (Y_n^0 / X) = \sum \theta_n^1 Y_n^1 / \sum_n \theta_n^0 Y_n^0. \tag{14}$$

Again it appears that there is a problem with monotonicity. For example, even if $Y_n^1 > Y_n^0$ for all j, it is possible for $\sum_n \theta_n^1 Y_n^1 < \sum_n \theta_n^0 Y_n^0$, and hence for there to be a decline in productivity growth as measured by (13).

Consider the numerical examples in table 1. For simplicity, inputs are assumed to be the same for each firm, and constant across periods so that productivity growth is determined by output growth. The examples show that output, and hence productivity, increases for both firms yet aggregate productivity growth, according to these methods, falls.

So, we have seen that the starting points for both of the currently dominant methods for doing decompositions of aggregate productivity have a serious problem. That is, the failure to satisfy the monotonicity property.

4 An Alternative Method

Starting from the same point as the methods of section 3 implies starting from the situation where the total-factor productivity scores have already been calculated. Later in this section, we consider the case of starting from the basic problem of aggregating over inputs and outputs in constructing a TFP growth index.

As observed in section 3, problems in aggregation arose through the shares that are used as weights being allowed to change. Consider the following alternatives. First, we weight the TFP changes using period 0 weights to form a "Laspeyres-type" index of productivity change between periods 0 and 1 ($\triangle TFP_{\mathcal{L}}^{0,1}$), for firms $n = 1, \cdots, N$:

$$\triangle TFP_{\mathcal{L}}^{0,1} = \exp\left[\sum_{n=1}^{N} \theta_n^0 (\ln TFP_n^1 - \ln TFP_n^0)\right]. \tag{15}$$

Alternatively, we weight the TFP changes using period 1 weights to form a "Paasche-type" index of productivity change, $\triangle TFP_{\mathcal{P}}^{0,1}$:

$$\triangle TFP_{\mathcal{P}}^{0,1} = \exp\left[\sum_{n=1}^{N} \theta_n^1 (\ln TFP_n^1 - \ln TFP_n^0)\right]. \tag{16}$$

As the choice between period 0 and period 1 shares is essentially arbitrary, it is also possible to use a Törnqvist aggregator function to aggregate the TFP scores, as follows:

$$\triangle TFP_T^{0,1} = \exp\left[\sum_{n=1}^{N}(1/2)(\theta_n^1+\theta_n^0)(\ln TFP_n^1 - \ln TFP_n^0)\right]. \quad (17)$$

Here the arithmetic mean of the shares in the two periods is used to weight (log) changes in TFP. This can be compared to equation (9), where TFP_n^1 and TFP_n^1 get different share weights. As we saw in section 3, this leads to monotonocity problems which are avoided by the use of a symmetric arithmetic mean over the shares for the two periods. Applying this aggregation technique to the numerical examples from table 1 yields an increase in TFP of 23% in example 1 and 13% in example 2. These are far more sensible estimates of aggregate productivity growth than those given by the methods of the previous section.[3]

Equation (17) has the very nice property of being additive (in logarithms), so that the weighted (log) productivity changes for each firm can simply be added. Alternatively, the aggregate can be decomposed into the contribution from each firm.

Unfortunately, (17) has a problem when examining issues relating to firm entry and exit over time. That is, zero values of TFP in one or other period cannot be handled well in (17), as the log of zero is indeterminate. Thus, it seems that this alternative is only of use if the same set of firms exists in each period, or if the analysis is restricted to this set of firms. Despite this problem, this method is very neat in the way that it aggregates over firms and/or industries. In the case of aggregating industry data over the short-run (when we can expect most industries to exist in each period), then this seems a very attractive method.[4]

We now consider the case of starting from the basic problem of aggregating over inputs and outputs in constructing a TFP growth index. Consider a Törnqvist (1936) index for

[3]However, it can be shown that the same kind of productivity paradox as observed in section 2 can occur using aggregation using this method.

[4]This method was used by Bartelsman and Gray (1996) for aggregating U.S. productivity from industry level (their "Divisa" method). They also used equation (15) and a variant of equation (16).

output of sector j, $j = 1, \ldots, J$, for goods $m \in M_j$, between periods 0 and 1:

$$Q_{Yj} = \exp\left[\sum_{m \in M_j} (1/2)(\theta^1_{jm} + \theta^0_{jm})(\ln q^1_{jm} - \ln q^0_{jm})\right], \tag{18}$$

where $\theta^t_{jm} = (p^t_{jm} q^t_{jm})/(p^t_j \cdot q^t_j)$ for $t = 0, 1$, and p^t_j and q^t_j are the price and quantity vectors for industry j, respectively. An aggregate (across sectors) index would be of the same form, but with $\sum M_j = M$ goods and $\hat{\theta}^t_{jm} = (p^t_{jm} q^t_{jm})/(p^t \cdot q^t)$, where p^t and q^t are the price and quantity vectors for the whole economy respectively. Denote such an index by Q_Y.

$$Q_Y = \prod_{j=1}^{J} \hat{Q}_{Yj}, \tag{19}$$

where

$$\hat{Q}_{Yj} = \exp\left[\sum_{m \in M_j} (1/2)(\hat{\theta}^1_{jm} + \hat{\theta}^0_{jm})(\ln q^1_{jm} - \ln q^0_{jm})\right]. \tag{20}$$

Thus, $\hat{Q}_{Yj}$ gives the contribution of sector j to aggregate output growth. Input indexes can be defined in a similar fashion, and denoted by Q_X, and other obvious notation. Then sector j's productivity growth is given by:

$$R_j = Q_{Yj}/Q_{Xj}. \tag{21}$$

Aggregate productivity growth is then given by:

$$\begin{aligned} R &= Q_Y/Q_X \\ &= \prod \hat{Q}_{Yj} / \prod \hat{Q}_{Xj} \\ &= \prod (\hat{Q}_{Yj}/\hat{Q}_{Xj}) \\ &= \prod \hat{R}_j, \end{aligned} \tag{22}$$

where $\hat{R}_j$ is the contribution of the j^{th} sector to aggregate productivity growth.

Note that $\hat{R}_j \neq R_j$, as the shares used to calculate each are different. However, the relationship between these two measures of productivity can be expressed as follows:

$$\hat{R}_j = (R_j)^{\vartheta_j}. \tag{23}$$

This allows us to write the following:

$$R = \prod_{j=1}^{J} (R_j)^{\vartheta_j}. \tag{24}$$

That is, by calculating R_j and $\hat{R}_j$, we can calculate the ϑ_j, which are the contribution of changing shares on aggregate productivity growth.

Similarly, we can decompose industry productivity growth into contributions from each firm in industry j, $n \in N_j$:

$$R_j = \prod_{n \in N_j} (R_n)^{\vartheta_n}. \tag{25}$$

With N firms in the whole economy:

$$R = \prod_{n=1}^{N} (R_n)^{\vartheta_n}. \tag{26}$$

Hence, we can get a very detailed decomposition of aggregate productivity growth, right down to the level of each firm's contribution. In addition, we start from the basic problem of calculating productivity, rather than starting from the point of aggregating productivity.

Unfortunately, this method has the same problem as our Törnqvist index in equation (17), in that firm entry and exit cannot be handled well. That is a firm that did not exist in period t would have a quantity of zero for each good in this period, and hence we would be dividing positive period $t+1$ values by zero in, e.g. equation (18). A similar problem exists for firms that exit the industry and so have zero quantities in period $t+1$.

However the method does have some strengths, as for (17). In addition, the Törnqvist-index method described here started from the point of aggregating over outputs to construct an output index, and similarly for inputs. This is another strength of the method, as it can easily handle multiple inputs and outputs.

The next section proposes a method with similarly nice aggregation properties as (17), but which gets around the problems that can arise when there is entry and exit of firms.

5 Another Alternative Method

In the previous section the main attention has been on the aggregation of productivity of sectors or industries. However, we saw in the previous section that an examination of changes in industry structure using data on firm entry/exit can lead to a failure of our proposed method. As the aggregation methods of section 3 were proposed primarily as a starting point for disaggregating aggregate productivity into contributions from changes in productivity *and* industry structure, we propose another method which can be used to examine these issues.

Our alternative method is based on the Bennet (1920) indicator (Diewert, 1998).[5] This method follows more closely the currently used methods, as will be shown, but without their main drawback. However, there will still be the choice of whether to use the share of output or the share of costs to weight productivity levels in aggregation.

To begin, consider using period 0 shares to construct an aggregate productivity-change index between periods 0 and 1, $\triangle TFP_L^{0,1}$, that is a function of the productivity levels $\triangle TFP_n^t$, for firms $n = 1, \cdots, N$:

$$\triangle TFP_L^{0,1} = \sum_{n=1}^{N} \theta_n^0 (TFP_n^1 - TFP_n^0). \tag{27}$$

This is like a fixed-base Laspeyres index, where only the first period's weight is used in aggregation. Alternatively, a Paasche-type, current-base index ($\triangle TFP_P^{0,1}$) could be defined as follows:

$$\triangle TFP_P^{0,1} = \sum_{n=1}^{N} \theta_n^1 (TFP_n^1 - TFP_n^0). \tag{28}$$

[5] Diewert (1998) used the terminology "indicator" to distinguish this kind of index in differences from the usual ratio type of index number.

The choice between $\triangle TFP_L^{0,1}$ and $\triangle TFP_P^{0,1}$ is essentially arbitrary. Hence, let $\triangle TFP_B^{0,1}$ denote an aggregate Bennet productivity-change indicator between periods 0 and 1, that is a function of the productivity levels TFP_n^t, for firms $n = 1, \cdots, N$, $t = 0, 1$:

$$\triangle TFP_B^{0,1} = \sum_{n=1}^{N} (1/2)(\theta_n^1 + \theta_n^0)(TFP_n^1 - TFP_n^0), \tag{29}$$

where θ_n^t denotes the share of firm n in total output (or costs) for period $t = 0, 1$. Note that this indicator is additive, and it uses the arithmetic average of the shares of the firms in periods 0 and 1 to weight the changes in productivity levels.[6] This kind of aggregator function has been shown to have some very nice properties, both in terms of the desirable axioms that it satisfies and in terms of an underlying economic justification in particular contexts (see Diewert, 1998).[7]

It is possible to also define an aggregate share-change indicator function, $\triangle S_B^{0,1}$, of a similar type:

$$\triangle S_B^{0,1} = \sum_{n=1}^{N} (1/2)(TFP_n^1 + TFP_n^0)(\theta_n^1 - \theta_n^0), \tag{30}$$

where share changes are weighted by the arithmetic mean of productivity levels in periods 0 and 1. Then, it is interesting to note the following:

$$\triangle TFP_A^{0,1} = \sum_{n=1}^{N} \theta_n^1 TFP_n^1 - \sum_{n} \theta_n^0 TFP_n^0 = \triangle TFP_B^{0,1} + \triangle S_B^{0,1}. \tag{31}$$

That is, the share-weighted productivity change index (in levels) that was examined in section 3 can be written as the sum of a Bennet productivity-change indicator and a Bennet share-change indicator. This makes it clear that $\triangle TFP_A^{0,1}$ is not only a measure of aggregate productivity change.

However, the following kind of decomposition has been proposed to determine the sources

[6] In this sense, it is very much like an additive version of the Fisher Ideal index, which is the geometric mean of Laspeyres (fixed-based) and Paasche (current-base) indexes. Alternatively, it can be thought of as an additive version of the Törnqvist index seen in the previous section.

[7] However, it can be shown that the same kind of paradoxical result as in section 2 can occur with this indicator used to aggregate productivity changes.

of "aggregate productivity change" defined in this manner (Griliches and Regev, 1995; Haltiwanger, 2000; Balk, 2001):

$$\begin{aligned}\triangle TFP_A^{0,1} &= \sum_{n\in C}(1/2)(\theta_n^1+\theta_n^0)(TFP_n^1-TFP_n^0)\\ &+ \sum_{n\in C}(1/2)(TFP_n^1+TFP_n^0)(\theta_n^1-\theta_n^0)\\ &+ \sum_{n\in E}\theta_n^1 TFP_n^1-\sum_{n\in X}\theta_n^0 TFP_n^0, \end{aligned} \tag{32}$$

where C denotes "continuing" firms that exist in both periods, E denotes firms that enter the industry in period 1, X denotes firms that have exited the industry after period 0, and shares sum to one in each period.[8]

The first term in (32) is interpreted as giving the productivity contribution from the continuing firms, second term gives the contribution of changing shares between the continuing firms, the third term gives the contribution from entering firms which the last term gives the contribution from exiting firms. Unfortunately, as noted above, these are in fact not contributions to aggregate productivity change, but to a combination of productivity and share changes.[9] This is what leads to the monotonicity problems observed in section 3. This decomposition may still be of interest, but clearly it should not be interpreted as it current is in the literature.

These interpretation problems are not inconsequential. For example, Baily, Bartelsman and Haltiwanger (2001) used $\triangle TFP_A^{0,1}$ to examine the behaviour of labour productivity over the business cycle. However, as is clear from the discussion above, $\triangle TFP_A^{0,1}$ is not a meaningful measure of productivity change for such an analysis. For example, due to the monotonicity problem, productivity in every firm could go up, yet the aggregate as measured by $\triangle TFP_A^{0,1}$ could go down. In addition, the Laspeyres quantity index discussed in section 3 is often used to construct real Gross Domestic Product (GDP), or some other output

[8]This decomposition can be changed in various minor ways in order to get different interpretations. See, e.g. Balk (2001; pp. 33–34).

[9]In the case of aggregating over firms (or industries) that exist in each period then the first term in (32) is the same as $\triangle TFP_B^{0,1}$ in (29), and the second term is the same as $\triangle S_B^{0,1}$ in (30).

aggregate.[10] The Laspeyres quantity index for aggregating over goods q_m^t, $m = 1, \cdots, M$ between time periods $t = 0, 1$, can be written as follows:

$$Q_L^{0,1} = \frac{\sum_m p_m^0 q_m^1}{\sum_m p_m^0 q_m^0} = \sum_m s_m^0 \left(\frac{q_m^1}{q_m^0} \right), \quad (33)$$

where s_m^0 is the value share of good m in the value of total output in period 0. That is, the change in quantities from period 0 to 1 are being weighted by a constant share, not a different share for each period. However, $\triangle TFP_A^{0,1}$ and $\triangle TFP_G^{0,1}$ use changing shares as weights. Hence, in looking for co-movements in a Laspeyres output index and TFP aggregates constructed using $\triangle TFP_A^{0,1}$ or $\triangle TFP_G^{0,1}$ seems inappropriate given that these TFP measures include share movements, but the quantity index does not. Hence, the cyclical properties of these TFP aggregates seem to be the wrong place to start such an analysis.

Hence, in order to examine these sorts of contributions to an actual indicator of aggregate productivity, we decompose the Bennet productivity indicator, $\triangle TFP_B^{0,1}$, from (29) into similar components as in (32).

$$\begin{aligned} \triangle TFP_B^{0,1} &= \sum_{n \in C} \theta_n^0 (TFP_n^1 - TFP_n^0) \\ &+ \sum_{n \in C} (1/2)(\theta_n^1 - \theta_n^0)(TFP_n^1 - TFP_n^0) \\ &+ \sum_{n \in E} (1/2)\theta_n^1 TFP_n^1 - \sum_{n \in X} (1/2)\theta_n^0 TFP_n^0. \end{aligned} \quad (34)$$

The first term is the contribution of productivity change from the continuing firms to the aggregate productivity change index, given that their shares in output remain unchanged from period 0. This ensures that share movements and productivity movements are not confounded. The last two terms give the contributions from entering firms and exiting firms

[10] The real output measure of Bartelsman and Gray (1996) is an implicit Laspeyres quantity index which they get by dividing the value of output by a Laspeyres price index. Baily, Bartelsman and Haltiwanger (2001) use this as their measure of real output.

respectively. This leaves the second line, which necessarily gives the contribution from the changing shares of the continuing firms.

One problem with this decomposition is that as presented it will not be invariant to the units of measurement. Hence, as is done empirically with equation (11) of section 3, it is possible to normalise (34) to put it in terms of growth rates. Here we see that we need at least three periods of data to make this possible, at least using the most obvious method. That is, to make $\triangle TFP_B^{1,2}$ invariant to the units of measurement, we need to divide by $\triangle TFP_B^{0,1}$, and so need data for periods $t = 0, 1, 2$. If there exist only two periods of data, then it is still possible to calculate contributions to aggregate productivity change, by dividing both sides of (34) by $\triangle TFP_B^{0,1}$, and hence get additive contributions to aggregate productivity change that sum to one.

An alternative is to normalise the productivity levels by one period's value for each firm. This creates a series of relative (to the selected base period) productivity levels. Applying $\triangle TFP_B^{0,1}$ to this series yields an aggregate indicator which is invariant to the units of measurement.

Finally, the alternative methods are compared in the examples in table 2, using data from the NBER Manufacturing Productivity Database of Bartelsman and Gray (1996). The current version of this database consists of information 450 four-digit manufacturing industries, 1958–1994, with the data mainly originating from the U.S. Census Bureau's Annual Survey of Manufactures and Census of Manufactures. The database includes estimates of TFP for each of the industries. Hence, the estimates are used to see if clear examples of problems in aggregation can be seen in these actual data.[11] Weights are taken to be the share in total employment. Adjacent SIC code industries were aggregated in pairs to see if there were any problems with monotonicity at this basic level.

Using $\triangle TFP_A^{0,1}$, there were 20 cases of positive TFP growth for each industry aggregating into negative TFP growth for the two-industry aggregate, and 17 cases of positive TFP growth for each industry aggregating into negative TFP growth. Two examples of these

[11] "TFP1" from the database is used in the following calculations. This is normalized to one in 1987 in the database.

cases are presented in table 2, with the corresponding results from using $\triangle TFP_G^{0,1}$, $\triangle TFP_T^{0,1}$ and $\triangle TFP_B^{0,1}$. In the first example, $\triangle TFP_G^{0,1} = 1.5\%$ and $\triangle TFP_A^{0,1} = 1.6\%$, although TFP for both industries was lower in 1992 than 1991. However, $\triangle TFP_T^{0,1} = -0.6\%$ and $\triangle TFP_T^{0,1} = -1.4\%$, which both have the appropriate sign. Similarly for the second example, except in this case TFP for both industries has risen, yet is measured to have fallen by the first two methods, while the alternative methods yield the appropriate sign.

6 Conclusion

It has been demonstrated that many problems can arise in aggregating productivity. In particular, a "productivity paradox" has been presented, which shows that when comparing two countries productivity growth may be higher in one country in all sectors than the others, yet it may have lower productivity growth overall. In addition, the two most commonly used methods for aggregating productivity in studies of the impact of structural change on aggregate productivity performance are shown to have a serious problem. That is, productivity for each firm may increase, yet overall productivity growth may decline.

An aggregation method was suggested which overcomes this "monotonicity" problem. This method has some very attractive aggregation and decomposition properties. However, while it is potentially very useful at the level of aggregating sectorial productivity, it is less useful in examining issues relating to the entry and exit of firms.

Hence, an alternative method is suggested which can overcome the aggregation problems inherent in the other methods, and can be used for examining changes in industry structure through the entry and exit of firms and changes in the relative shares of economic activity between firms.

References

Baily, M.N., E.J. Bartelsman and J. Haltiwanger (2001), "Labour Productivity: Structural Change and Cyclical Dynamics," *Review of Economics and Statistics* 83(3), 420–433.

Baily, M.N., C. Hulten and D. Campbell (1992), "Productivity Dynamics in Manufacturing Plants," *Brookings Papers: Microeconomics*, 187–267.

Balk, B.M. (2001), "The Residual: On Monitoring and Benchmarking Firms, Industries, and Economics with respect to Productivity," Inagugural Address, Erasmus Research Institute of Management, Erasmus University of Rotterdam.

Bartelsman, E.J. and W. Gray (1996), "The NBER Manufacturing Productivity Database," NBER Technical Working Paper 205.

Bennet, T.L. (1920), "The Theory of Measurement of Changes in Cost of Living," *Journal of the Royal Statistics Society* 83, 455–462.

Caves, D.W., L.R. Christensen and W. E. Diewert (1982a), "Multilateral Comparisons of Output, Input, and Producitivity Using Superlative Index Numbers," *The Economic Journal* 92, 73–86.

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

Diewert, W.E. (1992), "Fisher Ideal Output, Input, and Productivity Indexes Revisited," *Journal of Productivity Analysis* 3, 211–248.

Diewert, W.E. (1998), "Index Number Theory Using Differences Rather Than Ratios," Department of Economics Discussion Paper No. 98-10, University of British Columbia.

Diewert, W.E. (1999), "Axiomatic and Economic Approaches to International Comparisons," in A. Heston and R.E. Lipsey, eds., *International and Interarea Comparisons of Income, Output and Prices*, Chicago University Press, NBER, Chicago.

Domar, E.D. (1961), "On the Measurement of Technological Change," *The Economic Journal* LXXI, 709–729.

Griliches, Z. and H. Regev (1995), "Firm Productivity in Israeli Industry, 1979–1988," *Journal of Econometrics* 65, 175–203.

Gullickson, W. and M.J. Harper (1999), "Possible Measurement Bias in Aggregate Productivity Growth," *Monthly Labor Review*, February, 47–67.

Hahn, C-H. (2000), "Entry, Exit, and Aggregate Productivity Growth: Micro Evidence on Korean Manufacturing," Economics Department Working Papers No. 272, Organisation for Economic Co-operation and Development.

Haltiwanger, J. (2000), "Aggregate Growth: What Have we Learned from Microeconomic Evidence?" Economics Department Working Papers No. 267, Organisation for Economic Co-operation and Development.

Törnqvist, L. (1936), "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin*, vol. 10, 1–8.

Table 1: Numerical Examples of Monotonicity Problems

	y_1	y_2	$y_1 + y_2$	Share firm 1	Share firm 2	Geo.	Arith.
Example 1							
Period 1	1	19	20	0.05	0.95	16.4	18.1
Period 2	10	20	30	0.34	0.66	15.9	16.7
TFP Growth						-3.2%	-7.9%
Example 2							
Period 1	1	39	40	0.025	0.975	35.6	38.0
Period 2	10	40	50	0.2	0.8	30.3	34.0
TFP Growth						-14.8%	-10.6%

Notes: Inputs are assumed to be the same for each firm, and constant across periods so that productivity growth is determined by output growth. "Geo." refers to the share-weighted geometric-mean approach of aggregating TFP scores given in equation (9), while "Arith." refers to the share-weighted arithmetic-mean approach of aggregating TFP scores given in equation (11).

Table 2: Empirical Examples of Monotonicity Problems

	$TFP1$	$TFP2$	Share 1	Share 2	Geo.	Arith.	Törnqvist	Bennet
SICs 2393+2394								
1991	1.0344	0.8967	0.2578	0.7422	0.9304	0.9322		
1992	1.0296	0.8778	0.4559	0.5441	0.9440	0.9470		
TFP Growth					1.5%	1.6%	-0.6%	-1.4%
SICs 3769+3792								
1992	1.2574	1.0149	0.5309	0.4691	1.1372	1.1437		
1993	1.2672	1.0414	0.4301	0.5699	1.1331	1.1385		
TFP Growth					-0.4%	-0.5%	0.7%	1.8%

Notes: The SIC codes are as follows: 2303=Textile bags; 2394=Canvas and related products; 3769=Guided Missile and Space Vehicle Parts and Auxiliary Equipment, NEC; 3792=Travel trailers and campers. "Geo." refers to the share-weighted geometric-mean approach of aggregating TFP scores given in equation (9), while "Arith." refers to the share-weighted arithmetic-mean approach of aggregating TFP scores given in equation (11).