The Residual: On Monitoring and Benchmarking Firms, Industries, and Economies with Respect to Productivity

Bert M. Balk
Erasmus Research Institute of Management
Erasmus University Rotterdam
and
Methods and Informatics Department
Statistics Netherlands
P. O. Box 4000, 2270 JM Voorburg, The Netherlands
E-mail bblk@cbs.nl

Journal of Productivity Analysis 20(2003), 5-47

Abstract

Productivity is an important component of profitability, and therefore an important variable for monitoring and benchmarking exercises. This survey discusses the basic accounting model as well as the various measurement problems one gets involved in. By virtue of its structural features, this model is applicable to individual firms and aggregates such as industries or economies.

Though the measurement of productivity change and productivity differences is important, still more important is their explanation. Thus, first, this article reviews recent results relating to the decomposition of aggregate productivity change into components due to firm dynamics and intra-firm productivity change, results which were obtained by studying longitudinal enterprise microdata sets. Second, this article reviews a number of methods for decomposing productivity change and productivity differences, whether at the individual firm level or at aggregate level, into partial measures relating to technological change and efficiency change. The combination of both research strategies seems to be a promising undertaking.

Keywords: Producer behaviour; profitability; Total Factor Productivity; decomposition; firm level data; index number theory.

1 Introduction

There are two main dimensions in which the performance of, say, a firm can be assessed. The first is the dimension of time. The basic question here is: how is this or that firm doing over time? Assessing the performance of a firm over time is also called monitoring. The second dimension is characterized by the question: how is this or that firm doing relative to other, similar firms? To answer this question one needs to specify the reference set of firms and one needs sufficient information on each of the members of this set. This activity is usually called benchmarking. A combination of both dimensions is also possible. One is then said to be concerned with monitoring a set of firms over time.

The specific performance measure of course depends on the purpose of the exercise. In a market environment, however, a suitable overall performance measure seems to be profit, here defined as a firm's revenue minus its cost, or profitability, here defined as a firm's revenue divided by its cost. As will appear later on in this article, the profitability measure is better suited for intertemporal and interfirm comparisons than the profit measure.

An important component of profitability appears to be productivity. Indeed, as will be shown, the most encompassing measure of productivity change, Total Factor Productivity change, is nothing but the 'real' component of profitability change. Put otherwise, if there were no effect of prices then productivity change would coincide with profitability change. This is why productivity measurement in general, and monitoring and benchmarking firms with respect to productivity in particular, is so important.

The foregoing applies not only to individual firms but also to aggregates of firms, such as industries, industrial sectors, or even entire economies.[1]

[1]In the National Accounts aggregate profit occurs as net operating surplus at the cost

Traditionally, the monitoring of industries and economies is a task executed by national statistical agencies. The framework for performing this task is known as the System of National Accounts. The benchmarking of industries and economies, by making international comparisons, is a task executed by international organizations such as the OECD. But also a number of private organizations are active in this field. The interested parties are to be found among those responsible for economic policy, politicians, employers organizations, and labour unions. Measuring productivity levels and productivity change accurately is a necessary prerequisite for any policy directed at productivity growth, this being one of the determinants of welfare growth.[2]

Any measurement exercise must start with setting up an adequate accounting model. In such a model one must specify the inputs and the outputs, the quantities and the prices which must be observed, and the various concepts that play a role, such as revenue, cost, profit(ability), and value added. This will be the topic of section 2.

For ease of presentation, in this article mainly the vocabulary related to the dimension of time will be used. Thus, in section 3 we turn to the various instruments used for monitoring a firm. One can be interested in the development over time of a firm's revenue, its cost, its profit, or its value added. Most important, however, is the problem of decomposing any change into the contributions of price change and quantity change. Put otherwise, it is most important to split any nominal change into a monetary (price induced) part and a 'real' part. That is, one wants to be able to answer the question: how would revenue, cost, profit, or value added have changed in the absence of price changes? Thus, in this section we must review the basics of price and quantity index theory.

After all this, in a certain sense, preliminary work, section 4 turns to the productivity measures, to be used in comparisons over time as well as in inter-firm comparisons. The basic insight obtained here is that Total Factor Productivity change is the 'real' component of profitability change. But there are more productivity measures in use. They can be classified into

side of the industry and economy balances.

[2]Hulten (2001) concludes that "The residual [*i.e.*, TFP change] is a measure of the shift in the supply-side constraint on welfare improvement, but it is not intended as a direct measure of this improvement. To confuse the two is to confuse the constraint with the objective function." Basu and Fernald (2002), however, show that "productivity rather than, say, GDP or NNP, is the right measure of economic welfare under fairly general conditions."

two groups, according to the output concept used, and according to whether all input factors are taken into account or only a specific category of them (usually labour).

The section closes with two examples. The first is concerned with the U. S. economy over the past 50 years. The second is concerned with some 30 economies over the last 10 years. The first illustrates the famous slowdown of productivity, that started in the seventies, and its resurgence in the nineties. The second illustrates the large differences in performance, over time and between countries.

In the history of productivity measurement the attention was by and large focused at the level of aggregates. Section 5 presents a very condensed survey of the two main lines of research, the first directed at improving measurement, and the second directed at explanation. On the last line the concept of a 'representative firm' and the assumption that this firm always behaved optimally used to play an important role. This role came under attack when an increasing number of researchers got access to firm-level microdata. The perception of the inherent heterogeneity of reality and the often inefficient behaviour of firms has virtually terminated the 'representative firm' paradigm.

Thus, section 6 proceeds with the problem of how to decompose aggregate productivity change. Various factors appear to play a role: the coming and going of firms, the expansion or contraction of firms, and the productivity change at the individual firm level. The attention of researchers has clearly changed from explaining aggregate productivity change to explaining firm-level productivity change with help of suitable correlates. A number of recent empirical findings will be summarized.

In section 7 we go a step further and turn to the decomposition of productivity change itself. The old idea was that productivity change could be equated to technological change. This, however, appears to hold only in an economically perfect world. In reality there are a number of other factors contributing to productivity change, such as efficiency change, scale effects, and input- or output-mix change. The last 25 years have witnessed the development of a number of powerful techniques for measuring and decomposing productivity change at the individual firm level. These techniques can also easily be used for inter-firm comparisons and for time-series as well as cross-section analyses of non-market firms and institutions.

Section 8 concludes by pointing out some directions for further research.

2 The basic model

We consider a single production unit. This could be an establishment, a firm, an industry, or even an entire economy. For simplicity's sake, however, we will speak of a single firm and return to the issue of aggregation later on.

This firm will here be considered as an input-output system. At the output side we have the commodities produced: goods and/or services. Especially in the area of services it is not at all a trivial task to define precisely what the products of a firm are. Particularly difficult are financial institutions such as banks and insurance companies.

At the input side we have the various commodities – again: goods and services – consumed by the firm. Traditionally one distinguishes between a number of broad categories, which have intuitive appeal. First there is the group of reproducible capital inputs: buildings and other structures, machinery, tools. In short, everything that is not completely used up during the accounting period in which it was purchased, the accounting period usually being a year. Second, there are the various labour inputs: the work done by people of various age and education, part-time or full-time employees. Third, the energy used by the firm: gas, electricity, and water. Fourth, the materials used in the production process, which could be subdivided into raw materials, semi-fabricates, and auxiliary products. Fifth, and finally, the business services which are acquired for maintaining the production process. Again, it is not at all a trivial task to define precisely all the inputs and to classify them into these five categories.[3]

We will assume, however, that this can be done so that for the output side we have a list of commodities, which we will label with natural numbers $1, ..., M$, and for the input side a similar list, with labels $1, ..., N$. A commodity is a set of closely related items which, for the purpose of analysis, can be considered to be "equivalent", either in the static sense of their quantities being additive or in the dynamic sense of displaying proportional price or quantity changes.

Our next assumption is that this firm operates in a market environment,

[3] Traditionally the distinction was between capital, labour, and materials inputs. The oil crisis of the seventies led researchers to separate energy from materials, whereas the increasing importance of the service sector led them to separate business services also. Diewert (2000) suggested to separate leased capital from business services. He also suggested to consider as additional input categories: land, depletable resources, (monetary) working capital, knowledge capital, and infrastructure capital.

so that every commodity comes with a value (in monetary terms) and a price and/or a quantity. If value and price are available, then the quantity is obtained by dividing the value by the price. If value and quantity are available, then the price is obtained by dividing the value by the quantity. In any case, for every commodity it must be so that value equals price times quantity, the magnitudes of which of course pertain to the same agreed-on accounting period. Technically speaking, the price concept used here is the unit value.

All of this seems pretty trivial. The foregoing, however, hides a number of difficult and largely unresolved problems in economic measurement. We list here some of them:[4]

- With respect to capital assets it is well known that the calculation of their user costs and the split between their price and quantity components is a very data demanding task, the outcome of which moreover appears to depend on quite a number of assumptions. These include assumptions on the lifetime of the assets, the form of depreciation or asset efficiency, the reference interest rate, and the treatment of anticipated asset price changes. Also the utilization rate should be taken into account. See Hulten (1990) and Diewert (2001b) for (still) authoritative surveys of the statistical problems involved here and ways to tackle them.

- Production and consumption in the economic sense (sales, purchases) is often correlated with physical production and consumption. But not always. In the latter case, the question arises how to handle inventories of input or output commodities. This problem is especially important for firms involved in wholesale or retail trade. An interesting attempt to account for inventories at a distribution firm was developed by Diewert and Smith (1994).

- The production process often leads to the production of undesirable commodities. How do we handle these? Should, for instance, pollution be considered as an output or an input? And what (shadow) prices should be placed on environmentally undesirable commodities?

[4]A more detailed survey was provided by Diewert (2000). See Brynjolfsson and Hitt (2000) for additional topics.

- Some firms produce unique commodities, that is, commodities which are made on demand. Which accounting rules must then be followed?
- How must one value outputs whose production takes longer than the accounting period? Put otherwise, how to value work-in-progress?
- How to value the flow of services of intangible capital inputs, such as investments in software, organizational practices, new skills, or other forms of 'knowledge capital'?
- Especially problematic is the distinction between price and quantity of services. Services cannot be kept in stock and have frequently a unique character.

Assuming that, at least pragmatically, all these problems can be solved, it is now time to introduce some notation in order to define the various concepts we are going to use. As stated, at the output side we have M commodities, each with their price p_m^{it} and quantity y_m^{it}, where $m = 1, ..., M$, i is a firm label, and t denotes an accounting period. Similarly, at the input side we have N commodities, each with their price w_n^{it} and quantity x_n^{it}, where $n = 1, ..., N$. To avoid notational clutter, obvious vector notation will be used throughout. All prices are assumed to be positive and all quantities are assumed to be non-negative.

The firm i's revenue, that is, the value of its gross output, during the accounting period t is

$$p^{it} \cdot y^{it} \equiv \sum_{m=1}^{M} p_m^{it} y_m^{it}, \tag{1}$$

whereas its production cost is given by

$$w^{it} \cdot x^{it} \equiv \sum_{n=1}^{N} w_n^{it} x_n^{it}. \tag{2}$$

The firm's profit (disregarding tax on production) is then given by its revenue minus its cost, that is

$$p^{it} \cdot y^{it} - w^{it} \cdot x^{it}. \tag{3}$$

As we will shortly see, it is frequently more convenient to use the concept of profitability. The firm's profitability is defined as its revenue divided by its cost, that is

$$p^{it} \cdot y^{it} / w^{it} \cdot x^{it}. \tag{4}$$

Notice that profitability expressed as a percentage equals the ratio of profit to cost.

An important concept in economic accounting systems is value added. To define this, we must use additional notation. All the inputs are assumed to be allocatable to the five, mutually disjunct, categories mentioned earlier, namely capital (K), labour (L), energy (E), materials (M), and services (S). The entire input price and quantity vectors can then be partitioned as $w^{it} = (w_K^{it}, w_L^{it}, w_E^{it}, w_M^{it}, w_S^{it})$ and $x^{it} = (x_K^{it}, x_L^{it}, x_E^{it}, x_M^{it}, x_S^{it})$ respectively. The firm's value added (VA) is now defined as its revenue minus the costs of energy, materials, and services, that is

$$VA^{it} \equiv p^{it} \cdot y^{it} - w_E^{it} \cdot x_E^{it} - w_M^{it} \cdot x_M^{it} - w_S^{it} \cdot x_S^{it}. \tag{5}$$

Energy, materials and services together form the category of intermediate inputs, that is, inputs which are acquired from other firms or imported. The value added concept subtracts the total cost of intermediate inputs from the revenue obtained, and in doing so essentially conceives the firm as producing value added (that is, money) from the primary input categories capital and labour.[5]

Although gross output, y^{it}, is the natural output concept, the value added concept is important when we wish to aggregate single firms to larger entities. Gross output consists of deliveries to final demand and intermediate products. The split between these two output categories depends very much on the level of aggregation. Value added is immune to this problem. It enables one to compare (firms belonging to) different industries. From a welfare-theoretic point of view the value added concept is important since value added can be conceived as the income that serves to remunerate the factors capital and labour.

[5]Value added minus labour cost, $VA^{it} - w_L^{it} \cdot x_L^{it}$, could be called the firm's gross profit.

3 Instruments for monitoring and benchmarking

The notation employed in the previous section permits us to monitor a number of different firms over a number of different accounting periods. In order to economize on notation we will employ the following convention. When we are considering a single firm over time, we will drop the firm label superscript. When we are considering a set of firms during the same time period, we will drop the accounting period superscript.

What precisely do we want to see? In the intertemporal framework we want to see the evolution of revenue, cost, profit, or value added. In the cross-section framework we want to see the difference between firms with respect to revenue, cost, profit, or value added. In both frameworks the measures can be formulated in terms of ratios or in terms of differences. And, most important, we want to split any ratio or difference into a part due to prices and a part due to quantities. For example, when monitoring a single firm over time, we want to see whether its revenue change is caused by changed prices or by changed quantities. Or, in case of a comparison of two firms, we want to see whether their revenue difference is due to different prices or different quantities. Put otherwise, in either of these cases we want to see which part of the change or difference is 'monetary' (or price induced) and which part is 'real'.

In order to avoid the situation where the reader must continuously switch between the two frameworks, in the remainder of this article the discussion will mainly be cast in terms of intertemporal comparisons. Thus, we consider two periods, labelled $t = 0$ (which will be called the base period) and $t = 1$ (which will be called the comparison period).

Let us first consider ratio type measures. We want to decompose the revenue ratio into two parts,

$$\frac{p^1 \cdot y^1}{p^0 \cdot y^0} = P_o(p^1, y^1, p^0, y^0) Q_o(p^1, y^1, p^0, y^0), \tag{6}$$

of which the first part, $P_o(p^1, y^1, p^0, y^0)$, measures the effect of differing prices and the second part, $Q_o(p^1, y^1, p^0, y^0)$, measures the effect of differing quantities. The first part is called an output price index number. It is the outcome of a function $P_o(.)$, called a price index, operating on the output prices and quantities of both periods. The second part is called an output quantity index number. It is the outcome of a quantity index, that is a function $Q_o(.)$, also operating on the output prices and quantities of both periods.

The price index and the quantity index can both be conceived as functions which aggregate all the numerous prices and quantities respectively. This leads us to the concept of real output, which is defined as

$$\begin{aligned} Y^0 &\equiv p^0 \cdot y^0 \\ Y^1 &\equiv p^0 \cdot y^0 Q_o(p^1, y^1, p^0, y^0) = p^1 \cdot y^1 / P_o(p^1, y^1, p^0, y^0), \end{aligned} \tag{7}$$

where the equality in the second line is a simple restatement of expression (6). For the base period, real output is simply put equal to revenue. For the comparison period, real output is defined as base period revenue inflated by the quantity index number, or, equivalently, as comparison period revenue deflated by the price index number. Put otherwise, comparison period real output is comparison period revenue at the 'price level' of the base period. In a sense, the real output concept allows us to conceive the firm as producing a single money-metric output, namely deflated revenue, instead of the M different outputs. Notice, however, that this rests on the rather arbitrary normalization applied to the base period.[6]

It is useful to illustrate the foregoing with an example. If one specifies the output quantity index to be the Laspeyres index, that is $Q_o(p^1, y^1, p^0, y^0) = p^0 \cdot y^1 / p^0 \cdot y^0$, then comparison period real output is $Y^1 = p^0 \cdot y^1$. This means that all comparison period output quantities are valued at base period prices. The same result is obtained if one specifies the output price index to be the Paasche index, that is $P_o(p^1, y^1, p^0, y^0) = p^1 \cdot y^1 / p^0 \cdot y^1$.

Likewise, we want to decompose the cost ratio into two parts,

$$\frac{w^1 \cdot x^1}{w^0 \cdot x^0} = P_i(w^1, x^1, w^0, x^0) Q_i(w^1, x^1, w^0, x^0). \tag{8}$$

the first of which is an input price index number and the second an input quantity index number. Notice that the functional forms of the price and quantity indices used to get the decomposition of the revenue ratio, at the output side of the firm, might differ from the functional forms of the indices used to get the decomposition of the cost ratio, at the input side of the firm.

[6]Instead of normalizing with respect to one of the two time periods considered, one could of course normalize with respect to a third time period.

Given (8), real input can be defined as

$$\begin{aligned} X^0 &\equiv w^0 \cdot x^0 \\ X^1 &\equiv w^0 \cdot x^0 Q_i(w^1, x^1, w^0, x^0) = w^1 \cdot x^1 / P_i(w^1, x^1, w^0, x^0), \end{aligned} \quad (9)$$

where the equality in the second line is a simple restatement of expression (8). For the base period, real input is simply put equal to cost. For the comparison period, real input is defined as base period cost inflated by the input quantity index number, or, equivalently, as comparison period cost deflated by the input price index number. Put otherwise, comparison period real input is comparison period cost at the 'price level' of the base period. In a sense, the real input concept allows us to conceive the firm as consuming a single money-metric input, namely deflated cost, instead of the N different inputs. Again, one should notice the normalization that is involved here.

Using input price and quantity indices, the combined capital and labour cost ratio could be decomposed as

$$\frac{w_K^1 \cdot x_K^1 + w_L^1 \cdot x_L^1}{w_K^0 \cdot x_K^0 + w_L^0 \cdot x_L^0} = P_i(w_{KL}^1, x_{KL}^1, w_{KL}^0, x_{KL}^0) Q_i(w_{KL}^1, x_{KL}^1, w_{KL}^0, x_{KL}^0), \quad (10)$$

where $w_{KL}^t \equiv (w_K^t, w_L^t)$ and $x_{KL}^t \equiv (x_K^t, x_L^t)$ are the vectors of prices and quantities of the capital and labour inputs. Real capital and labour input is then defined as

$$\begin{aligned} X_{KL}^0 &\equiv w_K^0 \cdot x_K^0 + w_L^0 \cdot x_L^0 \\ X_{KL}^1 &\equiv (w_K^0 \cdot x_K^0 + w_L^0 \cdot x_L^0) Q_i(w_{KL}^1, x_{KL}^1, w_{KL}^0, x_{KL}^0) \\ &= (w_K^1 \cdot x_K^1 + w_L^1 \cdot x_L^1) / P_i(w_{KL}^1, x_{KL}^1, w_{KL}^0, x_{KL}^0). \end{aligned} \quad (11)$$

An important, frequently monitored, categorial cost ratio is the labour cost ratio, $w_L^1 \cdot x_L^1 / w_L^0 \cdot x_L^0$. Input price and quantity indices could be used to decompose this ratio as

$$\frac{w_L^1 \cdot x_L^1}{w_L^0 \cdot x_L^0} = P_i(w_L^1, x_L^1, w_L^0, x_L^0) Q_i(w_L^1, x_L^1, w_L^0, x_L^0). \quad (12)$$

The first part is a labour price index number, and the second part a labour quantity index number. These indices could of course be used to define the concept of real labour input, namely as

$$
\begin{aligned}
X_L^0 &\equiv w_L^0 \cdot x_L^0 \\
X_L^1 &\equiv w_L^0 \cdot x_L^0 Q_i(w_L^1, x_L^1, w_L^0, x_L^0) \\
&= w_L^1 \cdot x_L^1 / P_i(w_L^1, x_L^1, w_L^0, x_L^0).
\end{aligned} \tag{13}
$$

As defined in the previous section, profit is revenue minus cost. Provided that the base period profit is positive,

$$
\frac{p^1 \cdot y^1 - w^1 \cdot x^1}{p^0 \cdot y^0 - w^0 \cdot x^0} = \tag{14}
$$
$$
P_{io}(p^1, y^1, w^1, x^1, p^0, y^0, w^0, x^0) Q_{io}(p^1, y^1, w^1, x^1, p^0, y^0, w^0, x^0)
$$

would be the desired decomposition of the profit ratio. Since profit depends on inputs as well as outputs, we expect the price and quantity components of the profit ratio to depend on input as well as output variables. In particular, we expect these two components to depend on output and input price indices and output and input quantity indices respectively. Due to the simultaneous occurrence of a multiplicative and an additive element in the profit ratio, however, this relation is not a trivial one. The tool used to connect the decompositions (6), (8) and (14) is the logarithmic mean, for the definition of which the reader is referred to the Appendix.

Setting $\pi^t \equiv p^t \cdot y^t - w^t \cdot x^t$ $(t = 0, 1)$, we get by repeated application of the definition of the logarithmic mean

$$
\begin{aligned}
\ln\left(\frac{\pi^1}{\pi^0}\right) &= \frac{\pi^1 - \pi^0}{L(\pi^1, \pi^0)} = \\
&\frac{p^1 \cdot y^1 - w^1 \cdot x^1 - (p^0 \cdot y^0 - w^0 \cdot x^0)}{L(\pi^1, \pi^0)} = \\
&\frac{p^1 \cdot y^1 - p^0 \cdot y^0}{L(\pi^1, \pi^0)} - \frac{w^1 \cdot x^1 - w^0 \cdot x^0}{L(\pi^1, \pi^0)} = \\
&\frac{L(p^1 \cdot y^1, p^0 \cdot y^0) \ln(p^1 \cdot y^1 / p^0 \cdot y^0)}{L(\pi^1, \pi^0)} - \frac{L(w^1 \cdot x^1, w^0 \cdot x^0) \ln(w^1 \cdot x^1 / w^0 \cdot x^0)}{L(\pi^1, \pi^0)}.
\end{aligned} \tag{15}
$$

Using now expressions (6) and (8), the logarithm of the profit ratio can be expressed as

$$\ln\left(\frac{\pi^1}{\pi^0}\right) = \tag{16}$$
$$\frac{L(p^1 \cdot y^1, p^0 \cdot y^0)\ln(P_o(p^1,y^1,p^0,y^0)Q_o(p^1,y^1,p^0,y^0))}{L(\pi^1,\pi^0)} -$$
$$\frac{L(w^1 \cdot x^1, w^0 \cdot x^0)\ln(P_i(w^1,x^1,w^0,x^0)Q_i(w^1,x^1,w^0,x^0))}{L(\pi^1,\pi^0)}.$$

This can simply be rearranged to

$$\frac{\pi^1}{\pi^0} = \frac{P_o(p^1,y^1,p^0,y^0)^\phi}{P_i(w^1,x^1,w^0,x^0)^\psi}\frac{Q_o(p^1,y^1,p^0,y^0)^\phi}{Q_i(w^1,x^1,w^0,x^0)^\psi}, \tag{17}$$

where $\phi \equiv L(p^1 \cdot y^1, p^0 \cdot y^0)/L(\pi^1,\pi^0)$, that is, average revenue over average profit, and $\psi \equiv L(w^1 \cdot x^1, w^0 \cdot x^0)/L(\pi^1,\pi^0)$, that is, average cost over average profit. Connecting this result to (14), we see that the relation between the three price and quantity indices is given by

$$P_{io}(p^1,y^1,w^1,x^1,p^0,y^0,w^0,x^0) = \frac{P_o(p^1,y^1,p^0,y^0)^\phi}{P_i(w^1,x^1,w^0,x^0)^\psi} \tag{18}$$

$$Q_{io}(p^1,y^1,w^1,x^1,p^0,y^0,w^0,x^0) = \frac{Q_o(p^1,y^1,p^0,y^0)^\phi}{Q_i(w^1,x^1,w^0,x^0)^\psi}. \tag{19}$$

As we will see in the next section, the decomposition of the profitability ratio provides us with much simpler expressions.

The structure of value added is similar to that of profit. Thus, provided that the base period value added is positive, the desired decomposition of the value added ratio would be

$$\frac{VA^1}{VA^0} = \frac{p^1 \cdot y^1 - w_E^1 \cdot x_E^1 - w_M^1 \cdot x_M^1 - w_S^1 \cdot x_S^1}{p^0 \cdot y^0 - w_E^0 \cdot x_E^0 - w_M^0 \cdot x_M^0 - w_S^0 \cdot x_S^0} = \tag{20}$$
$$P_{io}(p^1,y^1,w_{EMS}^1,x_{EMS}^1,p^0,y^0,w_{EMS}^0,x_{EMS}^0) \times$$
$$Q_{io}(p^1,y^1,w_{EMS}^1,x_{EMS}^1,p^0,y^0,w_{EMS}^0,x_{EMS}^0),$$

where $w_{EMS}^t \equiv (w_E^t, w_M^t, w_S^t)$ and $x_{EMS}^t \equiv (x_E^t, x_M^t, x_S^t)$ are the vectors of prices and quantities of the intermediate inputs. With the first term at the right hand side of this expression we want to capture the contribution of changed prices, and with the second term we want to capture the contribution of changed quantities. It is straightforward to derive for the components of the value added ratio expressions similar to (18) and (19), whereby the price component is expressed as a function of the output price index and the three price indices of the intermediate input categories and the quantity component is expressed as a function of the output quantity index and the three quantity indices of the intermediate input categories.

Real value added (RVA) can then be defined as

$$\begin{aligned} RVA^0 &\equiv VA^0 \\ RVA^1 &\equiv VA^0 Q_{io}(p^1, y^1, w^1_{EMS}, x^1_{EMS}, p^0, y^0, w^0_{EMS}, x^0_{EMS}) \\ &= VA^1/P_{io}(p^1, y^1, w^1_{EMS}, x^1_{EMS}, p^0, y^0, w^0_{EMS}, x^0_{EMS}), \end{aligned} \tag{21}$$

that is, comparison period real value added is set equal to value added at the 'price level' of the base period. The concept of real value added allows us to conceive the firm as producing a single output, whose money-metric quantity at period t is given by RVA^t, from two categories of input, namely capital and labour.

We now briefly turn to difference type measures and additive decompositions. For example, an additive decomposition of the revenue difference would be

$$p^1 \cdot y^1 - p^0 \cdot y^0 = \mathcal{P}_o(p^1, y^1, p^0, y^0) + \mathcal{Q}_o(p^1, y^1, p^0, y^0), \tag{22}$$

of which the first term, $\mathcal{P}_o(p^1, y^1, p^0, y^0)$, measures the part of the revenue difference that is due to differing prices and the second term, $\mathcal{Q}_o(p^1, y^1, p^0, y^0)$, measures the part of the revenue difference that is due to differing quantities. The function $\mathcal{P}_o(.)$ is called an output price indicator and is assumed to have the prices and quantities of both periods as arguments. The function $\mathcal{Q}_o(.)$ is likewise called an output quantity indicator. Notice that both functions map price and quantity vectors into money amounts.

Using indicators, real output is defined as

$$\begin{aligned} Y^0 &\equiv p^0 \cdot y^0 \\ Y^1 &\equiv p^0 \cdot y^0 + \mathcal{Q}_o(p^1, y^1, p^0, y^0) = p^1 \cdot y^1 - \mathcal{P}_o(p^1, y^1, p^0, y^0), \end{aligned} \tag{23}$$

where the equality in the second line is a simple restatement of expression (22). It is useful to illustrate this with an example. If one specifies the output quantity indicator to be of the Laspeyres type, that is $\mathcal{Q}_o(p^1, y^1, p^0, y^0) = p^0 \cdot (y^1 - y^0)$, then comparison period real output is $Y^1 = p^0 \cdot y^1$. This means that all comparison period output quantities are valued at base period prices. The same result is obtained if one specifies the output price indicator to be of the Paasche type, that is $\mathcal{P}_o(p^1, y^1, p^0, y^0) = (p^1 - p^0) \cdot y^1$.

A decomposition of the cost difference would be

$$w^1 \cdot x^1 - w^0 \cdot x^0 = \mathcal{P}_i(w^1, x^1, w^0, x^0) + \mathcal{Q}_i(w^1, x^1, w^0, x^0), \tag{24}$$

where the first component measures the contribution of differing prices and the second the contribution of differing quantities. Using the functions $\mathcal{P}_i(.)$ and $\mathcal{Q}_i(.)$, the (combined capital and) labour cost difference could be decomposed similarly.

Since profit has by definition a linear structure, for the decomposition of the profit difference we can use the foregoing two equations to obtain

$$\begin{aligned}
&(p^1 \cdot y^1 - w^1 \cdot x^1) - (p^0 \cdot y^0 - w^0 \cdot x^0) = \\
&\quad (p^1 \cdot y^1 - p^0 \cdot y^0) - (w^1 \cdot x^1 - w^0 \cdot x^0) = \\
&\quad \mathcal{P}_o(p^1, y^1, p^0, y^0) + \mathcal{Q}_o(p^1, y^1, p^0, y^0) - \\
&\quad [\mathcal{P}_i(w^1, x^1, w^0, x^0) + \mathcal{Q}_i(w^1, x^1, w^0, x^0)] = \\
&\quad \mathcal{P}_o(p^1, y^1, p^0, y^0) - \mathcal{P}_i(w^1, x^1, w^0, x^0) + \\
&\quad \mathcal{Q}_o(p^1, y^1, p^0, y^0) - \mathcal{Q}_i(w^1, x^1, w^0, x^0).
\end{aligned} \tag{25}$$

The first two terms at the right hand side provide the price component, whereas the last two terms provide the quantity component of the profit difference. Thus, using difference type measures, there appears to be a very simple relation between the revenue and cost decompositions and the profit decomposition. A similar relation can easily be derived for the value added difference, $VA^1 - VA^0$.

It is useful to notice that, although ratio type measures and difference type measures can be developed independently, there appears to be a link in the sense that, provided that certain regularity conditions are met, every ratio type decomposition can be turned into a difference type decomposition and *vice versa*. The reader is referred to the Appendix for the mathematical details.

What are the advantages and disadvantages of ratio type measures *vis à vis* difference type measures? First of all, a ratio type measure is unitless and can simply be interpreted as 1 plus a percentage change. Difference type measures are advantageous in all situations where the magnitude that must be decomposed can take on values less than or equal to zero. Then a ratio type measure breaks down, either because dividing by zero is impossible or because the interpretation of a negative ratio or percentage is troublesome. Examples of magnitudes which can become less than zero are (price and quantity components of) profit and value added.

One can say that while economists usually prefer ratio type measures, business managers and accountants prefer difference type measures.

The important point now is: which formula should be selected as index or indicator? There are several theoretical approaches available, the most important of which are the axiomatic approach and the economic approach.

The axiomatic approach, with roots in the second half of the 19th century, specifies requirements which the formulas should satisfy. These requirements are called axioms or tests and are usually stated in the form of functional equations. The general idea is that an index or indicator is some kind of average of commodity specific changes. The basic theory for indices can be found in the monograph by Eichhorn and Voeller (1976) and the review article by Balk (1995). See also Diewert (1992). The parallel theory for indicators was developed by Diewert (1998).

The economic approach, with roots in the first half of the 20th century, defines theoretically motivated indices and indicators and combines assumptions on the behaviour of the firm (such as cost minimization, revenue maximization, or profit maximization) with assumptions on the prevailing production structure (formulated in terms of a production function, for instance) to obtain empirically applicable formulas. The basic theory for indices was surveyed by Balk (1998), and for indicators by Balk, Färe and Grosskopf (2001).

Although both approaches lead to a preference for certain specific formulas, it is fair to say that they do not lead to the recommendation of a single formula that could serve all imaginable purposes. If, in the axiomatic approach, the requirements are restricted to those that are more or less self-evident, then quite a number of formulas turn out to be satisfactory. On the other hand, every specific formula turns out to be characterized by at least one property which is not self-evident. With respect to the economic approach, it turns out that the assumptions needed to justify any specific formula are all more or less subject to argument. Put otherwise, available theory makes clear that the choice of a specific formula depends very much on the purpose one has in mind. Happily, however, it turns out that in the case of 'normal' and non-seasonal time series data all preferred formulas approximate each other reasonably well, at least when 'not too distant' time periods are compared.

Thus, more important than the theoretical problem of selecting the right formula are the many (practical) problems one encounters at the stage of implementation. In addition to those listed in section 2 we note the following problems:

- The data needed for calculating a preferred formula are not timely available, to the effect that a second-best formula must be used. The increasing availability of electronic transaction data, however, tends to mitigate this point somewhat.

- The universe of commodities at the input and output side of the firm is not constant but changes continuously. Put otherwise, one must deal with new and disappearing goods and services. In principle, these commodities do occur in the value figures of either of the periods which we wish to compare, but they become problematic when one proceeds to the task of decomposing ratios or differences of those figures.

- Many commodities, especially in the information and communication technology area, undergo a process of more or less rapid quality change. Just comparing quantities and nominal prices does not make much sense here. It is usually felt that quality change, whether improvement or deterioriation, belongs to the quantity component in a decomposition of revenue or cost change.

All this leads us to expect that actually calculated and published index numbers, whether by official agencies or by private organizations, will almost necessarily exhibit some degree of bias. The problems here are not unlike those in the field of the Consumer Price Index where the wellknown Boskin *et al.* (1996) commission report serves as a landmark. The recently completed Eurostat (2001) *Handbook on Price and Volume Measures in National Accounts*, where the production unit considered is an entire economy, can be considered as a research agenda. See also Diewert (2001a) for a list of research topics.

A prominent place on this research agenda is occupied by the problem of quantifying quality and variety change. Although over the years statistical agencies have acquired much experience here, and there is an extensive scientific literature, a number of theoretical and operational problems are still waiting for resolution. Much, but surely not enough, resources are being spent on the study of hedonic regression techniques. The operational worth of these techniques has for a long time be a topic of debate[7], but it seems that they are now gradually acquiring a recognized place in the day-to-day work of statistical agencies.[8] Jorgenson (2001) for example remarks that

> "The official [*i.e.*, U. S.] price indexes for computers and semiconductors provide the paradigm for economic measurement."

The huge literature on methods for dealing with quality and variety change will be surveyed in the forthcoming *CPI Manual* and *PPI Manual*, two joint publications by Eurostat, the International Labour Organization, the International Monetary Fund, the Organisation for Economic Co-operation and Development, the United Nations Economic Commission for Europe, and the World Bank.

4 Productivity measures

We are now in a position to discuss what to understand by 'productivity' and 'productivity change'. There appear to be several measures, the most important of which will be reviewed in this section.[9] The natural starting point is

[7]See Triplett (1990) for a review of (eight) reasons why statistical agencies have resisted hedonic methods. The adoption of hedonic methods was impeded by a number of conceptual issues, by doubts about the validity of the outcomes, and by the lack of suitable data.

[8]These techniques have in the meantime found their way into at least one academic textbook, namely that by Berndt (1991). Berndt and Rappaport (2001) provide a nice summary of work on desktop and mobile personal computers. The latest offspring on passenger cars is a study by Van Dalen and Bode (2002), an earlier version of which was presented at the Sixth Meeting of the International Working Group on Price Indices (Woolford 2001).

[9]This review follows to some extent the OECD (2001a) Manual. See also Schreyer and Pilat (2001).

to consider the ratio of comparison period and base period profitability, that is

$$\frac{p^1 \cdot y^1 / w^1 \cdot x^1}{p^0 \cdot y^0 / w^0 \cdot x^0}. \tag{26}$$

Using relations (6) and (8), this ratio can be decomposed as

$$\frac{p^1 \cdot y^1 / w^1 \cdot x^1}{p^0 \cdot y^0 / w^0 \cdot x^0} = \frac{p^1 \cdot y^1 / p^0 \cdot y^0}{w^1 \cdot x^1 / w^0 \cdot x^0} = \frac{P_o(p^1, y^1, p^0, y^0)}{P_i(w^1, x^1, w^0, x^0)} \frac{Q_o(p^1, y^1, p^0, y^0)}{Q_i(w^1, x^1, w^0, x^0)}. \tag{27}$$

The index of total factor productivity (TFP), for period 1 relative to period 0, is now defined by

$$ITFP^{10} \equiv \frac{Q_o(p^1, y^1, p^0, y^0)}{Q_i(w^1, x^1, w^0, x^0)}, \tag{28}$$

which is the real or quantity component of the profitability ratio. Put otherwise, $ITFP^{10}$ is the factor with which the output quantities on average have changed relative to the factor with which the input quantities on average have changed. If the ratio of these factors is larger (smaller) than 1, there is said to be productivity increase (decrease).

The wording used here suggests that a meaning can be attached to the term 'productivity' itself. Let us first consider the purely hypothetical situation of a firm which employs a single input to produce a single output. Then the index of TFP reduces to

$$ITFP^{10} = \frac{y^1/y^0}{x^1/x^0} = \frac{y^1/x^1}{y^0/x^0}, \tag{29}$$

which has indeed the simple interpretation as a ratio of productivities. In the single-input/single-output case y^t/x^t is the output quantity produced per unit of input quantity, which is a natural measure of the productivity of the production process. In the multi-input/multi-output case the term 'productivity' does not have such a natural sense.

Total factor productivity as a level concept can however be defined as

$$
\begin{aligned}
TFP^0 &\equiv p^0 \cdot y^0 / w^0 \cdot x^0 \\
TFP^1 &\equiv (p^0 \cdot y^0 / w^0 \cdot x^0) ITFP^{10}.
\end{aligned} \tag{30}
$$

Thus, base period TFP is set equal to base period profitability, and comparison period TFP is set equal to base period profitability multiplied by the index of TFP. Put otherwise, TFP could be called *real profitability*. Using the notation introduced in the previous section, we see that base period TFP can also be expressed as

$$
TFP^0 = Y^0 / X^0, \tag{31}
$$

and that, using again relations (6) and (8), comparison period TFP can be expressed as

$$
\begin{aligned}
TFP^1 &= \frac{p^0 \cdot y^0 Q_o(p^1, y^1, p^0, y^0)}{w^0 \cdot x^0 Q_i(w^1, x^1, w^0, x^0)} \\
&= \frac{p^1 \cdot y^1 / P_o(p^1, y^1, p^0, y^0)}{w^1 \cdot x^1 / P_i(w^1, x^1, w^0, x^0)} \\
&= Y^1 / X^1,
\end{aligned} \tag{32}
$$

that is, as real output divided by real input. This is in line with the single-input/single-output case. The relation between the index of TFP and the levels of TFP is now obviously given by

$$
ITFP^{10} = TFP^1 / TFP^0, \tag{33}
$$

but one should be aware of the normalization involved in defining the levels of TFP. The base period level is normalized as being equal to base period profitability.[10]

Using relation (27), the TFP index can also be expressed as

$$
ITFP^{10} = \frac{p^1 \cdot y^1 / w^1 \cdot x^1}{p^0 \cdot y^0 / w^0 \cdot x^0} \frac{P_i(w^1, x^1, w^0, x^0)}{P_o(p^1, y^1, p^0, y^0)}. \tag{34}
$$

[10] Thus one should be careful with interpreting real profitability as productivity in the case of a monopolist.

The right hand side of this expression consists of two parts. The first part is the profitability ratio. The second part is the ratio of an input price index number over an output price index number. Thus, if the profitability of the firm were not changing over time, then TFP change could be measured by the ratio of an input price index number over an output price index number. Put otherwise, if on average the input prices had increased more (less) than the output prices, then TFP change would be larger (smaller) than 1. Notice that constant profitability is not the same thing as constant profit.

The index of TFP takes into account all production factors, that is, all input categories. Traditionally, one speaks of a single factor productivity index when only one input category is taken into account.[11] Thus, for instance, the index of labour productivity is defined by

$$ILP^{10} \equiv \frac{Q_o(p^1, y^1, p^0, y^0)}{Q_i(w_L^1, x_L^1, w_L^0, x_L^0)}, \tag{35}$$

that is, the ratio of an output quantity index number over a labour input quantity index number. The corresponding level concept, labour productivity, is defined by Y^t/X_L^t, that is, real output divided by real labour input. By comparing their definitions, it is immediately clear that $ITFP^{10} = ILP^{10}$ if and only if $Q_i(w^1, x^1, w^0, x^0) = Q_i(w_L^1, x_L^1, w_L^0, x_L^0)$, that is, the average quantity change of all inputs is the same as the average quantity change of the labour inputs.

As noticed in the previous section, real value added is a frequently used output concept. The corresponding input categories are capital and labour. Thus, the index of value-added-based TFP is defined as the real value added ratio divided by the real capital and labour input ratio,

$$\begin{aligned} IVATFP^{10} &\equiv \frac{RVA^1/RVA^0}{X_{KL}^1/X_{KL}^0} \\ &= \frac{Q_{io}(p^1, y^1, w_{EMS}^1, x_{EMS}^1, p^0, y^0, w_{EMS}^0, x_{EMS}^0)}{Q_i(w_{KL}^1, x_{KL}^1, w_{KL}^0, x_{KL}^0)}, \end{aligned} \tag{36}$$

which is the ratio of a quantity index number of value added and a combined capital and labour input quantity index number. The corresponding level concept, that is value-added-based TFP, is defined by RVA^t/X_{KL}^t.

[11]Some authors speak of a multi factor productivity index when more than one input category is taken into account.

Similarly, the index of value-added-based labour productivity is defined by

$$IVALP^{10} \equiv \frac{Q_{io}(p^1, y^1, w^1_{EMS}, x^1_{EMS}, p^0, y^0, w^0_{EMS}, x^0_{EMS})}{Q_i(w^1_L, x^1_L, w^0_L, x^0_L)}, \tag{37}$$

which is the ratio of a quantity index number of value added and a labour input quantity index number. The corresponding level concept is defined by RVA^t/X^t_L. This is presumably the most widely used measure of productivity change, since it requires only readily available data. By comparing their definitions, it is clear that $IVATFP^{10} = IVALP^{10}$ if and only if $Q_i(w^1_{KL}, x^1_{KL}, w^0_{KL}, x^0_{KL}) = Q_i(w^1_L, x^1_L, w^0_L, x^0_L)$, that is, the average quantity change of the capital and labour inputs is the same as the average quantity change of the labour inputs.

The natural starting point for measuring productivity (change) in the difference framework is to consider the difference of comparison period and base period profit, $(p^1 \cdot y^1 - w^1 \cdot x^1) - (p^0 \cdot y^0 - w^0 \cdot x^0)$. Referring back to (25), TFP change is measured by

$$\Delta TFP^{10} \equiv \mathcal{Q}_o(p^1, y^1, p^0, y^0) - \mathcal{Q}_i(w^1, x^1, w^0, x^0), \tag{38}$$

which is an output quantity indicator minus an input quantity indicator. Notice that TFP change is now measured as an amount of money. An amount larger (smaller) than 0 indicates TFP increase (decrease). The elaboration of the other measures, as well as their interpretation, is straightforward and left to the reader.

Summarizing, there appear to be at least four different ways of measuring productivity change and productivity levels. The first main distinction is between total factor productivity and single factor productivity. The second main distinction is between using the 'natural', also called gross, output concept and the valued added output concept. Moreover, for each of these four alternatives there is a ratio and a difference type representation.

Productivity indexes or indicators are extremely important performance measures which can be used in a variety of circumstances. One can think of tracking the performance of a single firm or an aggregate of firms (an industry, or even the entire economy) over time; or, comparing the performance of a certain firm to similar firms, where similarity could be defined with respect to market or production technology; or, comparing the performance of a certain country's industry to the corresponding industries of other countries. The particular productivity measure that is thereby selected depends on the purpose of the exercise, the assumptions that can legitimately be made, and the availability of suitable data. For an in-depth discussion of the suitability of the various measures the reader is referred to the OECD (2001a) Manual. The TFP measure is, by definition, the most general measure of productivity change.[12]

Given the definition of the TFP index, expression (27) can be simplified to

$$\text{Profitability ratio } = ITFP^{10} \times \frac{P_o(p^1, y^1, p^0, y^0)}{P_i(w^1, x^1, w^0, x^0)}. \tag{39}$$

The term after the multiplication sign goes in the literature by different names: terms-of-trade index, price recovery index, or price performance index. The term measures the extent to which the average input price change is recovered by the average output price change. Thus, profitability change appears to be the combined result of TFP change and price performance. Moreover, by a slight redefinition of the two period labels, expression (39) could also be used to compare a firm's actual profitability to its targeted profitability.[13]

A regulation agency might use this expression as a vehicle for placing a bound on the average output price change by restricting a firm's profitability ratio to a prescribed value. Then the allowed rate of change of the output prices will be determined by the rate of change of the input prices corrected by the rate of TFP change. The last rate could be proxied by some industry- or economy-wide figure.

At the industry or economy level the labour productivity index appears to be a closely watched statistic, for instance in relation to wage negotiations. Recall that value added might be conceived as the remuneration that is to be

[12]The relation between the total factor productivity measures based on the two output concepts is discussed, in a production-theoretic framework, by Schreyer (2000). An empirical example is provided by Aldaz and Millán (2002).

[13]This expression strongly resembles the Profit Composition Analysis model developed by the New South Wales Treasury (1999) for analyzing the performance of regulated firms. The difference is that the PCA model starts with the profit difference instead of the profitability ratio, and concludes with an expression containing a mixture of ratio type and difference type measures. Decompositions like (39) occur frequently in the literature and go back at least to Miller (1984). A good survey was provided by Garrigosa and Grifell-Tatjé (1992).

distributed over the input factors capital and labour. A policy target could be to maintain the labour cost share in value added at some prescribed value, that is

$$\frac{w_L^0 \cdot x_L^0}{VA^0} = \frac{w_L^1 \cdot x_L^1}{VA^1}. \tag{40}$$

Using (12), (20) and (37), this policy target is seen to amount to

$$P_i(w_L^1, x_L^1, w_L^0, x_L^0) = IVALP^{10} P_{io}(p^1, y^1, w_{EMS}^1, x_{EMS}^1, p^0, y^0, w_{EMS}^0, x_{EMS}^0), \tag{41}$$

which means that the average price of labour input may rise by the (value-added-based) labour productivity index times the value added price index. The last index may be conceived as aggregating all price changes that are exogenous to the firm.

It is useful to close this section with two recent examples of applied work in this area. The first is a very instructive article by Jorgenson (2001). The production unit he considers is the U. S. economy. At its output side (Gross Domestic Product) he distinguishes between the following categories: investment goods, subdivided into non-IT, computers, software, and telecommunications equipment, and consumption goods, subdivided into non-IT goods and IT capital services. At the input side (Gross Domestic Income) he distinguishes between capital services, subdivided into non-IT, computers, software, and telecommunications equipment, and labour. His survey illuminates the challenging problems one encounters in obtaining meaningful price and quantity index numbers for all these commodity categories.

Table 1 presents some of Jorgenson's key results. The productivity slowdown, starting in the seventies, is clearly depicted as is the resurgence occurring in the second half of the nineties. Jorgenson concludes that this resurgence stems not only from the IT sectors of the economy but to an important degree also from the non-IT sectors. The explanation, however, appears to be still outstanding. Therefore, Jorgenson concludes that

> "Top priority must be given to identifying the impact of investment in IT at the industry level." and
>
> "The next priority is to trace the increase in aggregate TFP growth to its sources in individual industries."

Table 1: TFP change of the U. S. Economy

Period	Average yearly percentage
1948-1973	0.92
1973-1990	0.25
1990-1995	0.24
1995-1999	0.75

Source: Jorgenson (2001), Table 6.

Table 2: Labour productivity change

	Average yearly percentage	
	1990-1995	1995-2000
U. S. A.	0.8	2.6
E. U.	2.4	1.2
OECD	1.7	2.0
Netherlands	1.0	1.4

Source: McGuckin and Van Ark (2001), Table 2.

The second illustration is provided by a recent publication of The Conference Board (McGuckin and Van Ark 2001). This publication, entitled "Performance 2000: Productivity, Employment, and Income in the World's Economies", highlights the differences between some thirty economies over the last decade. This is an example of a comparison in a combined time series/cross-section (panel) framework. The additional layer of complexity is caused by the fact that prices not only change over time but also differ between the economies. Price differences between economies are captured by, what traditionally are called, purchasing power parities.[14]

The measure used in this publication is labour productivity, defined as GDP per hour worked. All value figures are converted with purchasing power parities to the U. S. 1996 price level. The differences in performance, summarized in Table 2, are striking. Again, the research question is, what is lying behind those aggregate figures?

[14]A recent survey of the theory of international price and quantity comparisons was provided by Balk (2001b).

5 Some history

Interesting details on the history of the concept of (total factor) productivity change can be found in Griliches (2001), the first section of which is a reworked version of his 1996 article on "the discovery of the residual." Another source of interesting historical details is Hulten (2001).

The first mention of TFP change as the ratio of an output quantity index and an input quantity index occurs in a contribution by Copeland (1937) in what, with hindsight, could be called the national income accounting approach. Stimulated by institutions such as the National Bureau of Economic Research, in the post-war period several studies were published, a typical one being Stigler (1947). These studies were mainly dealing with industry- or economy-wide aggregates. Although the TFP index was sometimes referred to as a measure of the efficiency of the economic process, the common opinion was best voiced by Abramowitz (1956), who called it a "measure of our ignorance."[15]

The other, production-theoretic approach appears to go back to Tinbergen (1942). He extended the Cobb-Douglas production function with a time trend variable. The difference between the growth rate of real output and a weighted average of the growth rates of real capital and labour input was interpreted variably as efficiency change, technical development, or "Rationalisierungsgeschwindigkeit".

The basic and very influential contribution of Solow (1957) can be conceived as some sort of linkage of both traditions. He showed that under certain conditions the parameters of the Cobb-Douglas production function could be equated to observable statistical magnitudes and the residual interpreted in terms of a ratio of output and input quantity index numbers. This is why the TFP index came to be known as the "Solow residual", although the name "residual" appears to have been used by Domar (1961) for the first time. Solow interpreted the residual as a measure of technical change.

Since the inception of the concept of TFP change there have been two main styles of research. The first was directed at explanation. The second was directed at better measurement, primarily of the input factors capital and labour. In the beginning, the second style was more prominent than the first. For example, Jorgenson and Griliches (1967) claimed that using the

[15]This has become a frequently repeated quote, the latest variation being Lipsey and Carlaw's (2000) conclusion that "TFP is as much a measure of our ignorance as it is a measure of anything positive."

"correct" index number framework and the "right" measurement of inputs would largely eliminate the role of the residual.

The residual disappeared indeed, but not at all due to better measurement techniques. The economy-wide disappearance of productivity growth in the seventies, its reappearance later on, and the search for the factors behind this world-wide phenomenon came to be known as the "productivity slowdown discussion". The emphasis shifted from measurement problems to explanation, and Griliches' work provides a clear demonstration of this shift. The main explanatory factors he considered were education, R&D expenditures, and patents.

The measurement problems, however, remained important. Looking back at a life-long of research in this area, Griliches (2001) says:

> "It is my hunch that at least part of what happened [namely, the productivity slowdown] is that the economy and its various technological thrusts moved into sectors and areas in which our measurement of output are especially poor: services, information activities, health, and also the underground economy."

but at the end of the day he concludes that

> "There have been many reasonable attempts to explain the productivity slowdown (...), but no smoking gun has been found, and no single explanation appears to be able to account for all the facts, leaving the field in an unsettled state until this day."

Until the nineties, the research on productivity change typically made use of the concept of the "representative firm" in combination with aggregate empirical material provided by statistical agencies. The increased availability of longitudinal enterprise microdata sets has opened up many new, exciting research possibilities.[16] Researchers are by now able to track large numbers of individual firms over time. This has led to a completely new area of research, with its own conferences[17] and research centers.[18]

[16]See for instance McGuckin (1995) or Heckman's (2001) Nobel Lecture.

[17]The international conferences on Comparative Analysis of Enterprise (micro)Data (Helsinki 1996, Bergamo 1997, The Hague 1999, Aarhus 2001) and the International Symposium on Linked Employer-Employee Data (Arlington VA 1999).

[18]These (usually confidential) microdata sets mainly originate from databases underlying aggregate figures published by national statistical agencies. They are, a.o., available for researchers at the Center for Economic Studies of the U. S. Bureau of the Census and the Center for Research of Economic Microdata (*Cerem*) of Statistics Netherlands.

6 Explaining aggregate productivity change

The explanation of aggregate productivity change, that is, productivity change at the level of an industry or an economy, starts with the truism that any aggregate is made up from a (large) number of individual firms. The relation between aggregate productivity change and firm-specific productivity change is, however, not a simple one. Though any aggregate can be conceived as a super-firm, and the same basic measurement model is applicable to aggregates and individual firms, such a super-firm is not the simple sum of a number of individual firms. In explaining aggregate productivity change one must not only deal with the temporal dynamics of the relevant population of firms, but also with the fact that these firms possibly interact with each other via transactions of goods and services. As will appear in this section, the dynamics has got a great deal of attention over the last years. The interaction, however, is a largely unexplored issue.[19]

Disregarding the interaction issue, the two main factors contributing to aggregate productivity change are *intra*-firm productivity change, and *inter*-firm reallocation. This reallocation is caused by the dynamic process of firm expansion, contraction, entry and exit. The first question, thus, is whether it is possible to distinguish unequivocally between all those factors.

As in the foregoing we will consider two periods. The set of firms existing at both periods will be denoted by C (continuing firms). The set of firms existing at the base period but vanished at the comparison period will be denoted by X (exiting firms), and the set of firms born after the base period and still existing at the comparison period will be denoted by N (entering firms). The productivity level of firm i at period t will be denoted by $PROD^{it}$. Each firm comes with some measure of relative size in the form of a weight θ^{it}. These weights add up to 1 for each period, that is

$$\sum_{i \in C \cup N} \theta^{i1} = \sum_{i \in C \cup X} \theta^{i0} = 1. \tag{42}$$

As shown in section 4, $PROD^{it}$ always has the form of real output (or real value added) divided by real input. Thus, ideally, the relative size measure θ^{it} should be consistent with the real input measure used in the denominator of $PROD^{it}$.

[19]Basic references on the relation between aggregate and individual measures of productivity change are Domar (1961), Hulten (1978), and Jorgenson (1980).

The aggregate productivity level at period t is then quite naturally defined as the weighted arithmetic average of the firm-specific productivity levels, that is $PROD^t \equiv \sum_i \theta^{it} PROD^{it}$, where the summation is taken over all firms existing at period t.[20] Aggregate productivity change between periods 0 and 1 is given by

$$PROD^1 - PROD^0 = \sum_{i \in C \cup N} \theta^{i1} PROD^{i1} - \sum_{i \in C \cup X} \theta^{i0} PROD^{i0}. \tag{43}$$

This can initially be decomposed as

$$\begin{aligned} & PROD^1 - PROD^0 = \\ & \quad \sum_{i \in N} \theta^{i1} PROD^{i1} \\ & \quad + \sum_{i \in C} \theta^{i1} PROD^{i1} - \sum_{i \in C} \theta^{i0} PROD^{i0} \\ & \quad - \sum_{i \in X} \theta^{i0} PROD^{i0}. \end{aligned} \tag{44}$$

The first term at the right hand side shows the contribution of entering firms, the second and third term together show the contribution of continuing firms, whereas the last term shows the contribution of exiting firms. The contribution of continuing firms is the outcome of the interaction between intra-firm productivity change, $PROD^{i1} - PROD^{i0}$, and inter-firm relative size change, $\theta^{i1} - \theta^{i0}$. There have been developed several methods to decompose this contribution further. We will review the various possibilities.

The first method decomposes the contribution of the continuing firms into a Laspeyres-type contribution of intra-firm productivity change and a Paasche-type contribution of relative size change:

$$\begin{aligned} & PROD^1 - PROD^0 = \\ & \quad \sum_{i \in N} \theta^{i1} PROD^{i1} \\ & \quad + \sum_{i \in C} \theta^{i0} (PROD^{i1} - PROD^{i0}) + \sum_{i \in C} (\theta^{i1} - \theta^{i0}) PROD^{i1} \\ & \quad - \sum_{i \in X} \theta^{i0} PROD^{i0}. \end{aligned} \tag{45}$$

[20]For computational reasons one sometimes prefers to define the aggregate productivity level as the weighted geometric average, $\ln PROD^t \equiv \sum_i \theta^{it} \ln PROD^{it}$.

The second term at the right hand side relates to intra-firm productivity change and uses base period weights. It is therefore called a Laspeyres-type measure. The third term relates to relative size change and is weighted by comparison period productivity levels. It is therefore called a Paasche-type measure. This decomposition was used in the study of Baily *et al.* (1992).

Due to the fact that the base period and comparison period weights add up to 1, we can insert an arbitrary scalar a, to obtain

$$\begin{aligned} &PROD^1 - PROD^0 = \\ &\quad \sum_{i \in N} \theta^{i1}(PROD^{i1} - a) \\ &\quad + \sum_{i \in C} \theta^{i0}(PROD^{i1} - PROD^{i0}) + \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i1} - a) \\ &\quad - \sum_{i \in X} \theta^{i0}(PROD^{i0} - a). \end{aligned} \tag{46}$$

Thus, entering firms contribute positively to aggregate productivity change insofar their comparison period productivity level exceeds a, and exiting firms contribute positively insofar their base period productivity level falls short of a. Since there are two different periods involved here, it is not quite clear what value for a it would be reasonable to take.

The second method uses a Paasche-type measure for intra-firm productivity change and a Laspeyres-type measure for relative size change. This leads to

$$\begin{aligned} &PROD^1 - PROD^0 = \\ &\quad \sum_{i \in N} \theta^{i1}(PROD^{i1} - a) \\ &\quad + \sum_{i \in C} \theta^{i1}(PROD^{i1} - PROD^{i0}) + \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i0} - a) \\ &\quad - \sum_{i \in X} \theta^{i0}(PROD^{i0} - a). \end{aligned} \tag{47}$$

It is possible to avoid the choice between the Laspeyres-Paasche-type and the Paasche-Laspeyres-type decomposition. The third method uses for the contribution of both intra-firm productivity change and relative size change Laspeyres-type measures. However, this simplicity is counterbalanced by the necessity to introduce a covariance-type term:

$$
\begin{aligned}
&PROD^1 - PROD^0 = \\
&\quad \sum_{i \in N} \theta^{i1}(PROD^{i1} - a) \\
&\quad + \sum_{i \in C} \theta^{i0}(PROD^{i1} - PROD^{i0}) + \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i0} - a) \\
&\quad + \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i1} - PROD^{i0}) \\
&\quad - \sum_{i \in X} \theta^{i0}(PROD^{i0} - a). \qquad (48)
\end{aligned}
$$

In view of the Laspeyres-type perspective, a natural choice for a now seems to be $PROD^0$, the base period aggregate productivity level. This leads to the decomposition advocated by Haltiwanger (1997).

Instead of using the Laspeyres perspective, one might however use the Paasche perspective. The covariance-type term accordingly appears with a negative sign. Thus, the fourth decomposition is

$$
\begin{aligned}
&PROD^1 - PROD^0 = \\
&\quad \sum_{i \in N} \theta^{i1}(PROD^{i1} - a) \\
&\quad + \sum_{i \in C} \theta^{i1}(PROD^{i1} - PROD^{i0}) + \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i1} - a) \\
&\quad - \sum_{i \in C}(\theta^{i1} - \theta^{i0})(PROD^{i1} - PROD^{i0}) \\
&\quad - \sum_{i \in X} \theta^{i0}(PROD^{i0} - a). \qquad (49)
\end{aligned}
$$

The natural choice for a would now be $PROD^1$, the comparison period aggregate productivity level.

The fifth method avoids the Laspeyres-Paasche dichotomy altogether, by using the symmetric method due to Bennet (1920). This amounts to taking the arithmetic average of the first and the second method. The covariance-type term then disappears. Thus,

$$PROD^1 - PROD^0 =$$

$$\begin{aligned}
&\sum_{i\in N}\theta^{i1}(PROD^{i1}-a)\\
&+(1/2)\sum_{i\in C}(\theta^{i1}+\theta^{i0})(PROD^{i1}-PROD^{i0})\\
&+(1/2)\sum_{i\in C}(\theta^{i1}-\theta^{i0})(PROD^{i1}+PROD^{i0}-2a)\\
&-\sum_{i\in X}\theta^{i0}(PROD^{i0}-a). \qquad (50)
\end{aligned}$$

A rather natural choice for a is now $(PROD^1 + PROD^0)/2$, the average aggregate productivity level. Substituting this in the last expression and rearranging somewhat, we finally get

$$\begin{aligned}
&PROD^1 - PROD^0 =\\
&\quad\sum_{i\in N}\theta^{i1}(PROD^{i1}-\frac{PROD^1+PROD^0}{2})\\
&\quad+\sum_{i\in C}\frac{\theta^{i1}+\theta^{i0}}{2}(PROD^{i1}-PROD^{i0})\\
&\quad+\sum_{i\in C}(\theta^{i1}-\theta^{i0})(\frac{PROD^{i1}+PROD^{i0}}{2}-\frac{PROD^1+PROD^0}{2})\\
&\quad-\sum_{i\in X}\theta^{i0}(PROD^{i0}-\frac{PROD^1+PROD^0}{2}). \qquad (51)
\end{aligned}$$

Thus, entering firms contribute positively to aggregate productivity change if their productivity level is above average. Similarly, exiting firms contribute positively if their productivity level is below average. Continuing firms can contribute positively in two ways: if their productivity level increases, or if the firms with above (below) average productivity levels increase (decrease) in relative size.[21] This decomposition is closely related to the one used by Griliches and Regev (1995). In view of its symmetry it should be the preferred one. Moreover, Haltiwanger (2000) notes that (51) is apt to be less sensitive to (random) measurement errors than (48).

This overview demonstrates a number of things. First, there is no unique decomposition of aggregate productivity change as defined by expression

[21] Notice that it can happen that for all continuing firms $PROD^{i1} > PROD^{i0}$ but that nevertheless the total contribution of the continuing firms to aggregate productivity change is negative. This 'paradox' is discussed by Fox (2002).

(43).[22] Second, one should be careful with reifying the different components, in particular the covariance-type term, since this term can be considered as being an artifact arising from the specific (Laspeyres- or Paasche-) perspective chosen. Third, the undetermined character of the scalar a lends additional arbitrariness to these decompositions. It is easily seen that letting a tend to 0 will lead to a larger contribution of the entering firms, the exiting firms, and the size change of continuing firms, at the expense of the intra-firm productivity change. Thus, it is to be expected that the outcome of any decomposition exercise will depend to some extent on the particular expression favoured by the researcher.

Having done with these, not unimportant, formalities an illustration is useful. The illustration draws on some results obtained by a team of national experts in an ongoing project of the Economics Department of the OECD. The novel feature of this project is that a common analytical framework was used on sets of longitudinal enterprise microdata from a number of member states. These data sets were, to the extent possible, harmonized. Most results obtained sofar are for total manufacturing. Table 3 presents some outcomes for aggregate labour productivity change. The decomposition method used is that of expression (51), whereby the shares are based on employment.

It appears that there are substantial differences between the annual percentage changes of aggregate labour productivity over the countries. This applies to both five yearly intervals. Further, entering and exiting firms appear to have a large influence. Sometimes the contributions of entry and exit go in the same direction, sometimes they go in opposite directions. Moreover, there appears to be a fair amount of reallocation between firms, the effect of which can go in either direction. However, by and large the intra-firm productivity change component tends to dominate the picture.[23]

The question thus shifts to the factors determining intra-firm productivity change. This has become an area of vigorous research, facilitated by the opportunities to link production survey type data to data coming from other kinds of firm level surveys, such as the European Community Innovation Surveys or the Wage Structure Surveys. There are some excellent reviews available which summarize the results obtained sofar: Bartelsman and Doms

[22] This non-uniqueness is a simple instance of the non-uniqueness experienced in so-called structural decomposition analysis, as widely used in input-output analysis; see Dietzenbacher and Los (1998).

[23] Limited information suggests that this is less so in the case of TFP.

Table 3: Decomposition of labour productivity change, total manufacturing

	Annual	Percentage share of each component			
	percentage	Entry	Within	Between	Exit
1985-1990					
Finland	5.4	0.4	72.5	7.0	20.1
France	2.0	-20.2	84.7	1.9	33.6
Italy	4.8	10.7	62.1	9.0	18.3
Netherlands	1.5	33.5	99.9	-8.1	-25.2
Portugal (1987-91)	6.6	-13.4	91.4	-9.7	31.8
United Kingdom	1.6	13.7	98.3	-7.4	-4.6
1990-1995					
Finland (1989-94)	4.6	-2.5	68.4	16.1	18.0
France	0.0				
W. Germany (1992-97)	2.1	-0.7	115.3	-12.1	-2.6
Italy	5.5	15.7	58.2	7.0	19.1
Netherlands	2.8	20.5	78.2	-10.8	12.1
Portugal	6.8	5.3	62.6	-4.3	36.4
United Kingdom (1987-93)	1.7	8.8	59.9	3.1	28.2

Source: OECD (2001b), Figure VII.1.

(2000), Haltiwanger (2000), and Ahn (2001), of which the last is the most comprehensive.

What are the main empirical findings? Bartelsman and Doms (2000) summarize the lessons as follows:

> "First, the amount of productivity dispersions is extremely large – some firms are substantially more productive than others. Second, highly productive firms today are more than likely to be highly productive firms tomorrow, although there is a fair amount of change in the productivity distribution. Third, a large portion of aggregate productivity growth is attributable to resource reallocation. The manufacturing sector is characterized by large shifts in employment and output across establishments every year – the aggregate data belie the tremendous amount of turmoil underneath. This turmoil is a major force contributing to productivity growth, resurrecting the Schumpeterian idea of creative-destruction. Fourth, quantifying the importance of various factors behind productivity growth, such as changes in the regu-

> latory environment or changes in technology, is a difficult task and has been only partially successful. Nonetheless, some useful lessons have been learned. In terms of the regulatory environment, any regulations that inhibit resource reallocation can have detrimental effects on productivity growth. Regarding the effect of technology on productivity, it is now known that documenting the correlation between a factor of production, such as computers, and productivity is not enough to understand causal mechanisms. Use of computers also is related to other variables correlated with productivity, such as human capital and managerial ability."

After reviewing quite a number of studies on productivity correlates such as regulation, management/ownership, technology and human capital, and international exposure, their conclusion is that

> "At the micro level, productivity remains very much a measure of our ignorance."

Ahn's (2001) conclusion is also worthwhile to quote here in full:

> "Both technology and human capital of workers appear to influence firm-level productivity. Innovative firms tend to shift the composition of their labour force toward more skilled labour through recruiting and training, and such shifts are often accompanied by higher productivity and higher wages for skilled labour.
>
> A direct causal link between technology or human capital and productivity at the individual level is difficult to prove, while evidence of technology-skill complementarity is widely observed. Both advanced technology use and higher wages may well be a result of a third factor (*e.g.* better management).
>
> Findings from micro data suggest that ownership structure is an important determinant of firm-level productivity. Likewise, exposure to competition, including international trade, plays a very important role in selecting high productivity firms.
>
> There are large and persistent differences in productivity levels across producers even in the same industry, and inputs and outputs are constantly reallocated from less efficient ones to more

efficient ones through firm dynamics. Aggregate productivity growth comes from firm dynamics as well as from within-firm productivity growth.

The contribution of firm dynamics to aggregate productivity appears to be more pronounced for total factor productivity growth than for labour productivity growth. While within-firm productivity growth seems to drive overall fluctuations in aggregate productivity growth, the contribution from the exit of low-productivity units increases its importance during cyclical downturns.

In spite of the large and still increasing share of the service sector in most OECD countries, difficulties in measuring service productivity have obliged most studies on firm dynamics and productivity growth to be focused on manufacturing. Emerging empirical studies suggest that firm dynamics are more volatile and more important for explaining aggregate productivity growth in the service sector than in the manufacturing sector."

The basic problem with measuring productivity change in the service sector is the unavailability of suitable price or quantity index numbers. It is therefore of utmost importance that statistical agencies try to close this gap.

7 What is productivity change?

As we have seen in the foregoing, several suggestions have been offered as an answer to the question: what is productivity change? In this section we will take a closer look at the meaning of productivity change at the individual firm level.

Measuring productivity change over time or comparing productivity levels between entities starts with positing something that is stable and/or communal. We will call this the technology and suppose that it is shared by at least the set of firms we wish to compare.

The classical approach was to represent the technology by a production function and to assume that all firms are behaving optimally in some economic sense, that is, for instance, as being profit-maximizers. The progress of the last two decades was brought about by recognizing the heterogeneity

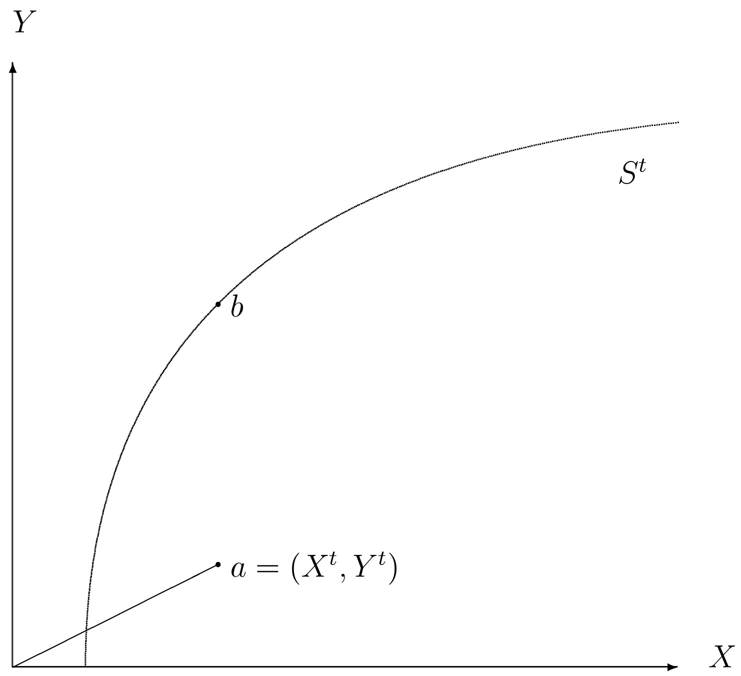

Figure 1: Total Factor Productivity

of reality, in the sense i) that the technology is a set rather than a function, and ii) that firms might behave non-optimally.

We will first illustrate the concept of TFP by a simple picture and then proceed to a discussion of the various factors which contribute to TFP change. We will thereby employ the various concepts defined in sections 3 and 4.

The horizontal axis in Figure 1 measures real input, whereas the vertical axis measures real output. Both are, as noticed earlier, conditional on a certain normalization with respect to input-mix and output-mix respectively. Put otherwise, the picture represents a single 'slice' of the full $N + M$-dimensional space of input and output quantities.

The technology of period t is to be thought of as the body of both tacit and explicit knowledge concerning products, processes, and organizational structures. Based on this body of knowledge there is a set of feasible combinations of input quantities and output quantities. In Figure 1 this set is represented by the area bounded by the curved line and the horizontal axis.

As depicted here, this set is assumed to exhibit some simple properties like free disposability of inputs and outputs. In reality, however, this set might have a less simple form.

The boundary of the technology set, that is the curved line itself, is called the frontier. This name is very appropriate, since beyond the frontier lie all those input-output combinations that are infeasible according to the technological state of affairs in period t. The mathematical representation of the frontier is the familiar production function $Y = F^t(X)$.

Each individual firm occupies a certain point within the technology set. Two examples have been drawn in the figure. The firm at point a uses real input X^t and produces real output Y^t. The TFP of this firm is then given by the ratio Y^t/X^t, which is just equal to the slope of the line connecting the origin O with the point a. Expanding real input X^t and real output Y^t with the same factor will leave TFP unchanged. Every other change in input or output quantities will in principle lead to TFP change. We will discuss now the various factors by which TFP can change.

As depicted, firm a is not particularly efficient. For instance, holding its real input X^t constant, the firm could expand its real output Y^t by a certain factor until it reaches the frontier. Or, holding its real output Y^t constant, it could contract its real input X^t by a certain factor until it reaches the frontier. Put otherwise, the firm can increase its efficiency by moving towards the frontier in the NW direction. This means that the slope of the line Oa increases, which is tantamount to saying that increasing efficiency means increasing TFP.

Consider now firm b. Since, as depicted, this firm is acting on the frontier, it is technically efficient. However, its TFP, that is the slope of the line Ob, can still change by moving on the frontier. There appear to be two logically distinct types of movement here:

1. The first is a movement within the 'slice' of the quantity space as drawn in the picture, that is a movement conditional on the firm's input- and output-mix. In particular, the firm could move towards the point where the slope of Ob attains its maximal value. This point would be reached when the line Ob became tangential to the frontier. At that point the firm's TFP would be maximal. This is what we will call the scale effect. The scale effect depends of course on the curvature of the frontier. Imagine, for instance, that the frontier is a straight line originating at O. Then a movement of firm b along this line would not change its TFP.

2. The firm can also move on the frontier by adapting its input- or output-

mix. This type of movement can of course not be represented in our simple figure since it cuts across all dimensions of the quantity space. Adaptation of the firm's input-mix can, for instance, be caused by a relaxation of capacity restrictions. Also, by moving towards the point where the firm is considered to be economically optimal, that is, the point where the firm, given the prices of all the inputs and outputs, maximizes profit, causes the input- or output-mix to change. At such a point the firm is called allocatively efficient.

Finally, the frontier itself can change over time. This means that the technology set changes, and is therefore called technological change.[24] An outwardbound change of the frontier is usually associated with technological progress, whereas an inwardbound change is associated with technological regress (which can occur as a result of organizational change). These changes can be of local nature, which means that a certain region can exhibit progress while an other region can exhibit regress. Assuming that our firm continues to stay on the frontier, technological change brings about TFP change.

It may be clear that, in order to arrive at measurement, all these rather intuitive notions must be made precise. The instruments needed in the first place are provided by duality theory.[25] Starting with the notion of a technology set S^t, duality theory shows that there are quite a number of equivalent representations of such a set in the form of mathematical functions. The main distinction thereby is between distance functions and value functions. Distance functions act on (primal) quantity space and are unitless. Value functions act on (dual) price space and read in money units. Well known among the distance functions are the (radial) input- and output distance functions. Well known among the value functions are the cost, revenue, and profit functions.

We will review the most important specifications, without introducing too much mathematical detail. For this, the reader is referred to the literature.[26]

We first discuss some output-orientated measures. The (direct) output distance function is most naturally defined by

[24]To be precise, this should be called *disembodied* technological change. Technological change as embodied in any input category is taken care of by the quality adjustment that must be made in order to make any 'new' input comparable to an 'old' input in quantity terms. See Lipsey and Carlaw (2000) for more on this issue.

[25]See Färe and Primont (1995) and Diewert (1982).

[26]See also the excellent, non-technical overview by Lovell (2000) with references to the more technical literature.

$$1/D_o^t(x,y) \equiv \sup\{\delta \mid \delta > 0, (x, \delta y) \in S^t\}. \quad (52)$$

The right hand side of this expression looks for the largest factor δ by which the output quantity vector y can be multiplied such that the resulting quantity vector δy is still producible by the input quantity vector x. The inverse of this largest factor is called the output distance function. This function is a (radial) measure of technical efficiency, which attains values between 0 and 1, conditional on a certain input quantity vector x and the output-mix implied by y.

The (direct) revenue function is defined by

$$R^t(x,p) \equiv \max_y\{p \cdot y \mid (x,y) \in S^t\}, \quad (53)$$

that is, the maximum revenue that can be obtained when output prices are given by p and the input quantities are fixed at x.

In some situations it is more appropriate to replace the condition of fixed input quantities by a budget constraint together with fixed input prices. Thus, the so-called indirect output distance function, defined by

$$1/ID_o^t(w/c,y) \equiv \sup\{\delta \mid \delta > 0, (x, \delta y) \in S^t, w \cdot x \leq c\}, \quad (54)$$

is again a measure of technical efficiency, based on the output-mix of y, but now conditional on the set of input quantity vectors which satisfy the requirement that the cost $w \cdot x$ does not exceed a given budget c. Likewise, the indirect revenue function is defined by

$$IR^t(w/c,p) \equiv \max_y\{p \cdot y \mid (x,y) \in S^t, w \cdot x \leq c\}, \quad (55)$$

that is, the maximum revenue that can be obtained when output prices are given by p and the input quantities are such that the cost at input prices w does not exceed the budget c.

For the input orientation a similar set of measures exist. The (direct) input distance function is most naturally defined by

$$1/D_i^t(x,y) \equiv \inf\{\delta \mid \delta > 0, (\delta x, y) \in S^t\}. \quad (56)$$

At the right hand side we now look for the smallest factor δ by which the input quantity vector x can be contracted such that δx is still able to produce the output quantity vector y. The inverse of this smallest factor is called the input distance function. The right hand side of the last expression itself is a measure of technical efficiency, conditional on the output quantity vector y and the input-mix given by x.

The (direct) cost function is defined by

$$C^t(w, y) \equiv \min_x \{w \cdot x \mid (x, y) \in S^t\}, \tag{57}$$

that is, the minimum cost that is necessary for producing the output quantities y when input prices are given by w.

The indirect functions replace the condition that output quantities be fixed by a revenue target together with an output price vector. Thus, the so-called indirect input distance function, defined by

$$1/ID_i^t(x, p/r) \equiv \inf\{\delta \mid \delta > 0, (\delta x, y) \in S^t, p \cdot y \geq r\}, \tag{58}$$

is again an inverse measure of technical efficiency based on x's input-mix, but now conditional on the set of output quantity vectors which satisfy the requirement that the revenue $p \cdot y$ attains at least a prescribed target r. Likewise, the indirect cost function is defined by

$$IC^t(w, p/r) \equiv \min_x \{w \cdot x \mid (x, y) \in S^t, p \cdot y \geq r\}, \tag{59}$$

that is, the minimum cost that is necessary, under input prices w, to yield revenue r when output prices are given by p.

Finally, the profit function is defined by

$$\Pi^t(w, p) \equiv \max_{x,y} \{p \cdot y - w \cdot x \mid (x, y) \in S^t\}, \tag{60}$$

that is, the maximum profit that can be obtained when output prices are p and input prices are w.

The fact that, without additional specifications, all these functions[27] represent the same technology enables the analyst to choose the analytical framework that fits 1) the behavioural objective that is assigned to or considered appropriate for the firms studied, and 2) the data available. For instance, suppose that the firms studied can be considered to be competitive profit maximizers, but that, for some reason, the analyst has only data on input

[27] In addition to the nine functions considered here, there are nonradial distance functions and various kinds of conditional distance and value functions.

prices and output quantities. Then an analysis in terms of the cost function is still appropriate, since profit maximization implies cost minimization.

By using these functions it is possible to replace the intuitive notions of technological change, technical efficiency change, allocative efficiency change, scale efficiency change, and input- or output-mix change by precisely formulated expressions which are adapted to the situation under study. Moreover, within the various frameworks it is possible to formulate hypotheses, for instance about the nature of technological change or about the scale properties of a technology.

The first question we now want to address is how these theoretical measures relate to the conventional, data-driven measures as discussed in section 4. This is among the main subjects of Balk's (1998) monograph. The results appear to be limited in scope.

One of the basic theoretical measures is what came to be called, due to Caves, Christensen and Diewert (1982), the (primal) Malmquist productivity index.[28] Depending on the situation studied, this index is defined as a function of (direct or indirect, input or output) distance functions. In the case of direct input distance functions the geometric average version reads

$$
\begin{aligned}
M_i(x^1, y^1, x^0, y^0) &\equiv \left[\frac{D_i^0(x^0, y^0)}{D_i^0(x^1, y^1)} \frac{D_i^1(x^0, y^0)}{D_i^1(x^1, y^1)}\right]^{1/2} \\
&= \frac{D_i^0(x^0, y^0)}{D_i^1(x^1, y^1)} \left[\frac{D_i^1(x^0, y^0)}{D_i^0(x^0, y^0)} \frac{D_i^1(x^1, y^1)}{D_i^0(x^1, y^1)}\right]^{1/2}. \qquad (61)
\end{aligned}
$$

The first part captures technical efficiency change (*i.e.*, the movement of the firm's position relative to the current frontier), whereas the second part captures technological change (*i.e.*, the movement of the frontier). Without knowledge of the distance functions, however, it is impossible to calculate $M_i(x^1, y^1, x^0, y^0)$ from the data.

Using various assumptions, it appears possible to relate this theoretical productivity index to an empirical index of the form (28). Specifically, one has to assume that the technology can be represented by a suitable[29]

[28] For the history related to this concept see Grosskopf (2002).

[29] The word 'suitable' could be used as a hyperlink to the whole body of theoretical results on flexible functional forms. Important contributions include Diewert and Wales (1987), (1988), (1992).

functional form which changes through time in a 'smooth' way; that the technology exhibits (locally) constant returns to scale; that the firm is and remains allocatively efficient, which means that, depending on the orientation chosen, its input-mix or output-mix is and remains optimal; that the firm, conditional on its input- or output-orientated technical efficiency, competitively maximizes profit. Under these assumptions it turns out that, depending on the specific functional form chosen, the Malmquist productivity index reduces either to the ratio of a Fisher output quantity index and a Fisher input quantity index or to the ratio of an (explicit or implicit) Törnqvist output quantity index and an (explicit or implicit) Törnqvist input quantity index.

Put otherwise, given all those assumptions, the TFP index appears to capture the combined effect of technological change and technical efficiency change. If one also were to assume that the firm is and remains technically efficient – which implies that the firm is cost efficient –, then the TFP index reduces to a measure of technological change. The whole set of assumptions leading up to this result – briefly summarized: a constant-returns-to-scale technology and a competitively profit-maximizing firm – reflects the classical position.

It may be clear that this position is not very realistic. Although one could argue that the assumption of constant returns to scale can validly be made on a global level and for the long run, it appears to be hardly tenable on a sectoral level and for the short run. And there is also sufficient evidence that firms are not behaving as nicely as theory would like them to do. However, any relaxation of assumptions comes at a price. We must invoke econometric methods in order to proceed.

Econometric methods are in the first place needed to estimate, within the framework chosen for the analysis, the function which represents the technology set S^t. Suppose that we have data (x^{it}, y^{it}) on firms $i = 1, ..., I$. There are a number of techniques available. The first we briefly consider is the method of activity analysis.

The basic idea of this method is that every pair (x^{it}, y^{it}) $(i = 1, ..., I)$ – that is, every observed activity – is an element of the set S^t. Thus S^t can be approximated by enveloping the observations as closely as possible – hence the alternative name Data Envelopment Analysis (DEA) – by piecewise linear contours. We consider two of those approximations. The first one,

$$S^t(CRS) \equiv \{(x,y) \mid \sum_{i'=1}^{I} z_{i'} x^{i't} \leq x, y \leq \sum_{i'=1}^{I} z_{i'} y^{i't}, \quad z_{i'} \geq 0 \; (i' = 1, ..., I)\}, \tag{62}$$

imposes global constant returns to scale. The second one, $S^t(VRS)$, which is defined by adding to the right hand side of (62) the constraint $\sum_{i'=1}^{I} z_{i'} = 1$, admits variable returns to scale. Since the addition of a restriction reduces the set of feasible elements, we have

$$S^t(VRS) \subseteq S^t(CRS), \tag{63}$$

that is, $S^t(VRS)$ envelops the data more closely than $S^t(CRS)$.

Based on these approximations, any input distance function value can be computed by solving the following linear programming problem

$$\begin{aligned} 1/D_i^t(x,y) = & \min_{z,\delta} \delta \text{ subject to} \\ & \sum_{i'=1}^{I} z_{i'} x^{i't} \leq \delta x, y \leq \sum_{i'=1}^{I} z_{i'} y^{i't}, \\ & z_{i'} \geq 0 \; (i' = 1, ..., I), [\sum_{i'=1}^{I} z_{i'} = 1], \end{aligned} \tag{64}$$

and any cost function value can be computed by solving the following linear programming problem

$$\begin{aligned} C^t(w,y) = & \min_{z,x} w \cdot x \text{ subject to} \\ & \sum_{i'=1}^{I} z_{i'} x^{i't} \leq x, y \leq \sum_{i'=1}^{I} z_{i'} y^{i't}, \\ & z_{i'} \geq 0 \; (i' = 1, ..., I), [\sum_{i'=1}^{I} z_{i'} = 1]. \end{aligned} \tag{65}$$

The restriction between brackets in these two equations must of course be deleted in the case of imposing global constant returns to scale.

For the other functions reviewed above similar linear programming problems could be stated. I refer to Färe, Grosskopf and Lovell (1994) for a detailed exposition of the theory. A more recent source is Cooper, Seiford and Tone (1999). There have been developed a number of (semi-) commercial software packages, such as Warwick DEA Software (see `www.deazone.com`), Frontier Analyst (see `www.banxia.com`), and On Front (see `www.emq.com`) to execute the necessary calculations.

The second technique is called stochastic frontier analysis (SFA). The basic idea behind this technique, or rather this set of techniques, can most easily be grasped by first considering the conventional approach.

Suppose that the firms under study could be considered as competitive cost minimizers, facing in each period the same prices. This means that, for each firm i $(i = 1, ..., I)$, its actual cost $c^{it} \equiv w^t \cdot x^{it}$ is equal to the minimum cost as given by the cost function, $C^t(w^t, y^{it})$. Since the actual form of the cost function is unknown, $C^t(w, y)$ must be replaced by a suitable functional form $f(w, y, t; \Phi)$, where Φ denotes a set of unknown parameters. A stochastic noise term is added, and Φ is to be estimated from a set of equations like

$$\ln c^{it} = \ln f(w^t, y^{it}, t; \Phi) + v^{it} \; (i = 1, ..., I). \tag{66}$$

The stochastic noise term is thereby usually assumed to be independent and identically distributed according to a normal distribution with mean zero.

Stochastic frontier analysis explicitly recognizes the fact that firms might not behave optimally. In the present example this means that actual cost c^{it} may be higher than minimum cost $C^t(w^t, y^{it})$, but never can be lower. This can be modelled by introducing an additional, asymmetrically distributed term, and replacing (66) by

$$\ln c^{it} = \ln f(w^t, y^{it}, t; \Phi) + u^{it} + v^{it} \; (i = 1, ..., I). \tag{67}$$

In this system of equations, as before, v^{it} represents noise and is accordingly distributed symmetrically around zero. But u^{it} represents inefficiency, is always non-negative, and must therefore follow an asymmetrical distribution (usually a truncated-at-zero normal distribution). Both terms are assumed to be independently distributed.

The foregoing paragraphs were only intended to give an idea of the approach pursued by stochastic frontier analysts.[30] The main features distinguishing SFA from DEA might, however, be clear. Whereas SFA is basically a regression method, yields a smooth frontier, is stochastic, and parametric, DEA is based on solving linear programming problems, yields a piecewise linear frontier, is deterministic, and nonparametric. Since its inception, a quarter of a century ago, the body of theory and applications relating to SFA has grown almost exponentially. The state of the art was recently reviewed by Kumbhakar and Lovell (2000). Coelli (1996) developed a non-commercial software package for stochastic frontier estimation.

The final approach considered here consists in specifying a complete parametric model. Again assuming that the cost function framework is the appropriate one, this approach starts off at what Balk (1997) called "the canonical form of cost function and cost share equations." The basic idea can be presented as follows.

Provided that some regularity conditions are met, firm i's actual cost c^{it} will satisfy the following relation

$$c^{it} ITE^{it} = C^t(w^{it*}, y^{it}), \tag{68}$$

where $ITE^{it} \equiv 1/D_i^t(x^{it}, y^{it})$ is the firm's input technical efficiency, $C^t(w, y)$ is the period t cost function, y^{it} is the firm's actual vector of output quantities, and w^{it*} is a vector of so-called shadow input prices. These shadow prices, which although as yet unknown can be proven to exist, serve to make the firm's actual cost as corrected by the firm's technical efficiency (which has, as we know, a value between 0 and 1) to be equal to the minimum cost as given by the cost function. Due to Shephard's Lemma, equation (68) can be supplemented by N equations relating the actual cost shares of the inputs to first-order derivatives of the cost function.[31]

The next step is to select a suitable functional form for the cost function. Since the cost function is time-dependent, this implies that some hypothesis on the nature of technological change must necessarily be incorporated. Next, in order to reduce the number of free parameters to a manageable size, one

[30] Fuentes, Grifell-Tatjé and Perelman (2001) provide an example where an output distance function is estimated.

[31] Since this system uses shadow input prices, it is sometimes referred to as a 'shadow cost function system', a term which is slightly misleading because it suggests that there is a different kind of (cost) function involved.

must model the firm-specific input technical efficiencies as well as the relation between the firm-specific shadow input prices and the actual sector-specific prices which the firms are facing. After all this work has been done, the resulting system of equations for costs and cost shares can be estimated by a suitable econometric method. For further details the reader is referred to Balk and Van Leeuwen (1999) and Balk (1998; section 8.3).

Once armed with an estimated version of some functional representation of the technology set S^t it becomes possible to compute the measures which can be defined for the various components of productivity change. For instance, the Malmquist index can be computed as well as its decomposition into technological change and technical efficiency change components. But one can also enhance the Malmquist index with components referring to scale efficiency change and input- or output-mix change. An example was recently provided by Balk (2001a).[32]

The framework sketched above can also be used for cross-section type comparisons of firms. Of course, in this setting there is no correlate to technological change since all firms in the comparison are supposed to share the same technology. But one can compare firms with respect to their technical efficiency, their scale efficiency, and their allocative efficiency. This is called *benchmarking*.

Moreover, this framework can be used for intertemporal and cross-sectional studies of non-market firms and similar institutions, such as hospitals, schools, prisons, and police districts. All one has to do is to select the functional representation for the technology that fits the data and that is considered to be an appropriate behavioural objective. A nice collection of such studies is to be found in the volume edited by Blank (2000). A more recent example is provided by Grosskopf and Moutray (2001). They used the Malmquist productivity index, based on the indirect output distance function, which was estimated by DEA, to measure the performance of public high schools over time. This article is also a nice illustration of the fact that the construction of appropriate input and output variables is not at all a trivial task.

[32] As appears from this article, there is some debate on how to measure the various components and how to relate those to the Malmquist index. Recent contributions include Zofio (2001) and Lovell (2001).

8 Conclusion

Measuring productivity change or productivity differences requires both good theory and good data. In the first sections of this survey I laid out the basic accounting model that ties together the various concepts which play a role. A basic insight offered in section 4 was that the natural measure of productivity change, gross output based Total Factor Productivity change, is the 'real' component of profitability change. Its computation, whether at aggregate or firm level, requires the splitting of all value changes at the output and input side of the production unit considered into price and quantity components. This is not at all a trivial task, as shown extensively in sections 2 and 3. Thus, along the route several topics for research and development have been indicated.

Until a decade or so ago, both measurement and explanation of productivity change or productivity differences was confined to the levels of aggregation as typically used by official statistical agencies. The advent of longitudinal enterprise microdata has lead many a researcher to get more insight into the turmoil hidden underneath aggregate figures.

As shown in section 6, much has been learned about the incredible dynamics of firms and the contribution of intra- and inter-firm factors to aggregate productivity change. However, firm-level productivity change as such remained more or less a black box. The logical step forward would therefore be to enhance this type of analysis by a decomposition of firm-level productivity change, using the methodology reviewed in section 7. A recent example, where the Malmquist productivity index together with its components technological change and technical efficiency change was computed for German manufacturing sector microdata over the period 1981-1993, was provided by Cantner and Hanusch (2001). This type of research could lead to a deeper insight into the evolutionary processes that are taking place within modern economies.

Such insight is not only important for its own sake but also for any government policy that aims at aggregate productivity growth. For the fine-tuning of such a policy some understanding of the various factors that alone or together contribute to productivity change is indispensable. This point was recently made by Diewert (2001c). Should economic policy be directed at pushing the technological frontiers ahead? Or should economic policy be directed at removing the barriers for (more) efficient behaviour?

As the example of economies of scale demonstrates, an even more refined form of analysis is called for. According to Diewert (2001c), (internal) economies of scale can be due to (1) the existence of indivisibilities, (2) the existence of fixed costs, (3) certain laws of geometry or physics, or (4) certain laws of probability. Each of these sources would require a separate approach. At this level the role of statistical figures for guiding economic policy must be taken over by carefully designed case studies, whose role it is to stimulate the imagination of all involved. It occurs to me that this is the traditional area of interest of business administration.

Appendix: Additive and multiplicative decompositions

In order to see the equivalence of additive and multiplicative decompositions, we make use of the simple but powerful tool called the logarithmic mean. This mean is, for two positive numbers a and b, defined by $L(a, b) \equiv (a-b)/\ln(a/b)$ and $L(a, a) \equiv a$. It is easy to check that the function $L(.)$ has all of the properties one expects a symmetric mean to possess. The logarithmic mean allows us to switch between a difference and a ratio.[33]

Thus, starting for instance with the multiplicative decomposition of the revenue ratio (6) we take the logarithm at both sides, so that we get

$$\ln\left(\frac{p^1 \cdot y^1}{p^0 \cdot y^0}\right) = \ln P_o(p^1, y^1, p^0, y^0) + \ln Q_o(p^1, y^1, p^0, y^0), \quad (69)$$

which can be written, using the definition of the logarithmic mean, as

$$\frac{p^1 \cdot y^1 - p^0 \cdot y^0}{L(p^1 \cdot y^1, p^0 \cdot y^0)} = \ln P_o(p^1, y^1, p^0, y^0) + \ln Q_o(p^1, y^1, p^0, y^0). \quad (70)$$

But this can be rearranged as

$$p^1 \cdot y^1 - p^0 \cdot y^0 = \quad (71)$$
$$L(p^1 \cdot y^1, p^0 \cdot y^0) \ln P_o(p^1, y^1, p^0, y^0) + L(p^1 \cdot y^1, p^0 \cdot y^0) \ln Q_o(p^1, y^1, p^0, y^0),$$

which is an additive decomposition of the revenue difference into a price indicator and a quantity indicator. Recall that $L(p^1 \cdot y^1, p^0 \cdot y^0)$ is an average of the period 1 revenue $p^1 \cdot y^1$ and the period 0 revenue $p^0 \cdot y^0$, and notice that $\ln P_o(.)$ and $\ln Q_o(.)$ are approximately equal to the percentage price and quantity change respectively.

Reversely, starting with an additive decomposition of the revenue difference (22), we can apply the logarithmic mean to get

$$L(p^1 \cdot y^1, p^0 \cdot y^0) \ln\left(\frac{p^1 \cdot y^1}{p^0 \cdot y^0}\right) = \mathcal{P}_o(p^1, y^1, p^0, y^0) + \mathcal{Q}_o(p^1, y^1, p^0, y^0). \quad (72)$$

[33] The logarithmic mean was introduced in the economics literature by Törnqvist in 1935 in an unpublished memo of the Bank of Finland; see Törnqvist, Vartia and Vartia (1985). It has the following properties: (1) $\min(a, b) \leq L(a, b) \leq \max(a, b)$; (2) $L(a, b)$ is continuous; (3) $L(\lambda a, \lambda b) = \lambda L(a, b)$ $(\lambda > 0)$; (4) $L(a, b) = L(b, a)$. A simple proof of the fact that $(ab)^{1/2} \leq L(a, b) \leq (a + b)/2$ was provided by Lorenzen (1990).

This can be rearranged as

$$\ln\left(\frac{p^1 \cdot y^1}{p^0 \cdot y^0}\right) = \frac{\mathcal{P}_o(p^1, y^1, p^0, y^0)}{L(p^1 \cdot y^1, p^0 \cdot y^0)} + \frac{\mathcal{Q}_o(p^1, y^1, p^0, y^0)}{L(p^1 \cdot y^1, p^0 \cdot y^0)}, \tag{73}$$

and further as

$$\frac{p^1 \cdot y^1}{p^0 \cdot y^0} = \exp\left(\frac{\mathcal{P}_o(p^1, y^1, p^0, y^0)}{L(p^1 \cdot y^1, p^0 \cdot y^0)}\right) \exp\left(\frac{\mathcal{Q}_o(p^1, y^1, p^0, y^0)}{L(p^1 \cdot y^1, p^0 \cdot y^0)}\right), \tag{74}$$

which clearly is a multiplicative decomposition of the revenue ratio into a price index number and a quantity index number.

Acknowledgments

The views expressed in this article are those of the author and do not necessarily reflect the policies of Statistics Netherlands. This article is a revised version of the inaugural lecture delivered at the occasion of accepting a chair at the Rotterdam School of Management, Erasmus University Rotterdam. The author thanks Knox Lovell for his invitation to submit the lecture to this Journal. He also thanks Paul de Boer, Erwin Diewert, Knox Lovell, as well as an unknown referee for many helpful remarks on the lecture version.

References

Abramowitz, M. (1956). "Resource and Output Trends in the U. S. since 1870." *The American Economic Review* 46, 5-23.

Ahn, S. (2001). "Firm Dynamics and Productivity Growth: A Review of Micro Evidence from OECD Countries." Working Paper No. 297, Economics Department, OECD, Paris.

Aldaz, N. and J. A. Millán. (2002). "Intermediate Inputs and Manufacturing Sectors Growth in the Spanish Regions." *Journal of Regional Science*, forthcoming.

Balk, B. M. (1995). "Axiomatic Price Index Theory: A Survey." *International Statistical Review* 63, 69-93.

Balk, B. M. (1997). "The Decomposition of Cost Efficiency and the Canonical Form of Cost Function and Cost Share Equations." *Economics Letters* 55, 45-51.

Balk, B. M. (1998). *Industrial Price, Quantity, and Productivity Indices: The Micro-Economic Theory and an Application.* Boston/Dordrecht/London: Kluwer Academic Publishers.

Balk, B. M. (2001a). "Scale Efficiency and Productivity Change." *Journal of Productivity Analysis* 15, 159-183.

Balk, B. M. (2001b). "Aggregation Methods in International Comparisons: What Have We Learned?" Report Series Research in Management, Erasmus Research Institute of Management, Erasmus University Rotterdam.

Balk, B. M., R. Färe, and S. Grosskopf. (2001). "The Theory of Economic Price and Quantity Indicators." Mimeo, Methods and Informatics Department, Statistics Netherlands, Voorburg.

Balk, B. M. and G. van Leeuwen. (1999). "Parametric Estimation of Technical and Allocative Efficiencies and Productivity Changes: A Case Study." In *Micro- and Macrodata of Firms: Statistical Analysis and International Comparison*, edited by S. Biffignandi. Heidelberg: Physica-Verlag.

Baily, M., C. Hulten, and D. Campbell. (1992). "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity: Microeconomics* 2, 187-249.

Bartelsman, E. J. and M. Doms. (2000). "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature* XXXVIII, 569-594.

Basu, S. and J. G. Fernald. (2002). "Aggregate Productivity and Aggregate Technology." *European Economic Review* 46, 963-991.

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

Berndt, E. R. (1991). *The Practice of Econometrics, Classic and Contemporary.* Reading MA: Addison-Wesley.

Berndt, E. R. and N. J. Rappaport. (2001). "Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview." *American Economic Association Papers and Proceedings, The American Economic Review* 91, 268-273.

Blank, J. L. T., editor. (2000). *Public Provision and Performance.* Amsterdam etc.: Elsevier.

Boskin, M. J., E. R. Dulberger, R. J. Gordon, Z. Griliches, and D. W. Jorgenson. (1996). "Toward a More Accurate Measure of the Cost of Living." Final Report to the [U. S.] Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index.

Brynjolfsson, E. and L. M. Hitt. (2000). "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives* 14(4), 23-48.

Cantner, U. and H. Hanusch. (2001). "Heterogeneity and Evolutionary Change: Empirical Conception, Findings and Unresolved Issues." In *Frontiers of Evolutionary Economics*, edited by J. Foster and J. S. Metcalfe. Cheltenham UK / Northampton MA: Edward Elgar.

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

Coelli, T. J. (1996). "A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation." CEPA Working Papers No. 7/96, Centre for Efficiency and Productivity Analysis, School of Economics, University of New England, Armidale NSW, Australia.

Cooper, W. W., L. M. Seiford, and K. Tone. (1999). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Boston / Dordrecht / London: Kluwer Academic Publishers.

Copeland, M. A. (1937). "Concepts of National Income." In *Studies in Income and Wealth, Volume 1.* New York: National Bureau of Economic Research.

Dietzenbacher, E. and B. Los. (1998). "Structural Decomposition Techniques: Sense and Sensitivity." *Economic Systems Research* 10, 307-323.

Diewert, W. E. (1982). "Duality Approaches to Microeconomic Theory." In *Handbook of Mathematical Economics, Volume II*, edited by K. J. Arrow and M. D. Intriligator. Amsterdam: North-Holland. Also in *Essays in Index Number Theory, Volume 1*, edited by W. E. Diewert and A. O. Nakamura. Amsterdam: North-Holland, 1993.

Diewert, W. E. (1992). "Fisher Ideal Output, Input, and Productivity Indexes Revisited." *The Journal of Productivity Analysis* 3, 211-248. Also in *Essays in Index Number Theory, Volume 1*, edited by W. E. Diewert and A. O. Nakamura. Amsterdam: North-Holland, 1993.

Diewert, W. E. (1998). "Index Number Theory using Differences rather than Ratios." Discussion Paper No. 98-10, Department of Economics, The University of British Columbia, Vancouver.

Diewert, W. E. (2000). "The Challenge of Total Factor Productivity Measurement." *International Productivity Monitor* 1, 45-52.

Diewert, W. E. (2001a). "Research in Price Measurement for the Next Twenty Years." Discussion Paper No. 01-11, Department of Economics, The University of British Columbia, Vancouver.

Diewert, W. E. (2001b). "Measuring the Price and Quantity of Capital Services under Alternative Assumptions." Discussion Paper No. 01-24, Department of Economics, The University of British Columbia, Vancouver.

Diewert, W. E. (2001c). "Notes on the Role of Government: To Facilitate Growth or to Provide Services?" Mimeo, Department of Economics, The University of British Columbia, Vancouver.

Diewert, W. E. and A. M. Smith. (1994). "Productivity Measurement for a Distribution Firm." *Journal of Productivity Analysis* 5, 335-347.

Diewert, W. E. and T. J. Wales. (1987). "Flexible Functional Forms and Global Curvature Conditions." *Econometrica* 55, 43-68.

Diewert, W. E. and T. J. Wales. (1988). "A Normalized Quadratic Semiflexible Functional Form." *Journal of Econometrics* 37, 327-342.

Diewert, W. E. and T. J. Wales. (1992). "Quadratic Spline Models for Producer's Supply and Demand Functions." *International Economic Review* 33, 705-722.

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

Eichhorn, W. and J. Voeller. (1976). *Theory of the Price Index.* Lecture Notes in Economics and Mathematical Systems 140, Berlin etc.: Springer-Verlag.

Eurostat. (2001). *Handbook on Price and Volume Measures in National Accounts.* Luxembourg: Office for Official Publications of the European Communities.

Färe, R., S. Grosskopf, and C. A. K. Lovell. (1994). *Production Frontiers.* Cambridge UK: Cambridge University Press.

Färe, R. and D. Primont. (1995). *Multi-Output Production and Duality: Theory and Applications.* Boston / London / Dordrecht: Kluwer Academic Publishers.

Fox, K. J. (2002). "Problems with (Dis)Aggregating Productivity, and Another Productivity Paradox." Draft, School of Economics, The University of New South Wales, Sydney.

Fuentes, H. J., E. Grifell-Tatjé, and S. Perelman. (2001). "A Parametric Distance Function Approach for Malmquist Productivity Index Estimation." *Journal of Productivity Analysis* 15, 79-94.

Garrigosa, E. G. and E. Grifell-Tatjé. (1992). "Profits and Total Factor Productivity: A Comparative Analysis." *Omega International Journal of Management Science* 20, 553-568.

Griliches, Z. (1996). "The Discovery of the Residual: A Historical Note." *Journal of Economic Literature* XXXIV, 1324-1330.

Griliches, Z. (2001). *R&D, Education, and Productivity: A Retrospective.* Cambridge MA: Harvard University Press.

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

Grosskopf, S. (2002). "Some Remarks on Productivity and its Decompositions." Mimeo. Department of Economics, Oregon State University, Corvallis OR.

Grosskopf, S. and C. Moutray. (2001). "Evaluating Performance in Chicago Public High Schools in the Wake of Decentralization." *Economics of Education Review* 20, 1-14.

Haltiwanger, J. (1997). "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence." *Federal Reserve Bank of St. Louis Economic Review* 79, No. 3, 55-77.

Haltiwanger, J. (2000). "Aggregate Growth: What have we Learned from Microeconomic Evidence?" Working Paper No. 267, Economics Department, OECD, Paris.

Heckman, J. J. (2001). "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109, 673-748.

Hulten, C. R. (1978). "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45, 511-518.

Hulten, C. R. (1990). "The Measurement of Capital." In *Fifty Years of Economic Measurement*, edited by E. R. Berndt and J. E. Triplett. Studies in Income and Wealth Volume 54, Chicago and London: The University of Chicago Press.

Hulten, C. R. (2001). "Total Factor Productivity: A Short Biography." In *New Developments in Productivity Analysis*, edited by C. R. Hulten, E. R. Dean and M. J. Harper. Studies in Income and Wealth Volume 63, Chicago and London: The University of Chicago Press.

Jorgenson, D. W. (1980). "Accounting for Capital." In *Capital, Efficiency and Growth*, edited by G. von Furstenberg. Cambridge UK: Ballinger.

Jorgenson, D. W. (2001). "Information Technology and the U. S. Economy." *The American Economic Review* 91, 1-32.

Jorgenson, D. W. and Z. Griliches. (1967). "The Explanation of Productivity Change." *Review of Economic Studies* 34, 249-283.

Kumbhakar, S. C., and C. A. K. Lovell. (2000). *Stochastic Frontier Analysis.* Cambridge UK: Cambridge University Press.

Lipsey, R. G. and K. Carlaw. (2000). "What does Total Factor Productivity Measure?" *International Productivity Monitor* 1, 31-40.

Lorenzen, G. (1990). "Konsistent Addierbare Relative Änderungen." *Allgemeines Statistisches Archiv* 74, 336-344.

Lovell, C. A. K. (2000). "Measuring Efficiency in the Public Sector." In *Public Provision and Performance*, edited by J. L. T. Blank. Amsterdam etc.: Elsevier.

Lovell, C. A. K. (2001). "The Decomposition of Malmquist Productivity Indexes." Mimeo, Department of Economics, University of Georgia, Athens GA.

McGuckin, R. H. (1995). "Establishment Microdata for Economic Research and Policy Analysis: Looking Beyond the Aggregates." *Journal of Business & Economic Statistics* 13, 121-126.

McGuckin, R. H. and B. van Ark. (2001). *Performance 2000: Productivity, Employment, and Income in the World's Economies.* New York: The Conference Board.

Miller, D. M. (1984). "Profitability = Productivity + Price Recovery." *Harvard Business Review* 62, 145-153.

New South Wales Treasury. (1999). "Profit Composition Analysis: A Technique for Linking Productivity Measurement & Financial Performance." Research & Information Paper TRP 99-5, Office of Financial Management, NSW Treasury, Sydney.

OECD. (2001a). *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth.* Paris: Organisation for Economic Cooperation and Development.

OECD. (2001b). "Chapter VII. Productivity and Firm Dynamics: Evidence from Microdata." *OECD Economic Outlook* No. 69.

Schreyer, P. (2000). "Separability, Path-Dependence and Value-Added Based Productivity Measures: An Attempt of Clarification." Mimeo, National Accounts Division, OECD, Paris.

Schreyer, P. and D. Pilat. (2001). "Measuring Productivity." *OECD Economic Studies* No. 33, 127-170.

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

Stigler, G. (1947). *Trends in Output and Employment.* New York: National Bureau of Economic Research.

Tinbergen, J. (1942). "Zur Theorie der Langfristigen Wirtschaftsentwicklung." *Weltwirtschafliches Archiv* 55, 511-549.

Törnqvist, L., P. Vartia, and Y. O. Vartia. (1985). "How Should Relative Changes Be Measured?" *The American Statistician* 39, 43-46.

Triplett, J. E. (1990). "Hedonic Methods in Statistical Agencies: An Intellectual Biopsy." In *Fifty Years of Economic Measurement*, edited by E. R. Berndt and J. E. Triplett. Studies in Income and Wealth Volume 54, Chicago and London: The University of Chicago Press.

Van Dalen, J. and B. Bode. (2002). "Quality-Corrected Price Indices: The Case of the Dutch New Passenger Car Market, 1990-1999." Mimeo, Rotterdam School of Management, Erasmus University Rotterdam.

Woolford, K., editor. (2001). *Papers and Proceedings of the Sixth Meeting of the International Working Group on Price Indices.* Canberra: Australian Bureau of Statistics.

Zofio, J. L. (2001). "Malmquist Productivity Index Decompositions: A Unifying Framework." Mimeo, Departamento de Análisis Económico, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain.