the review of income and wealth

Review of Income and Wealth
Series 58, Number 2, June 2012
DOI: 10.1111/j.1475-4991.2012.00504.x

# OUTPUT, OUTCOME, AND QUALITY ADJUSTMENT IN MEASURING HEALTH AND EDUCATION SERVICES

BY PAUL SCHREYER*

*OECD Statistics Directorate, Paris*

This paper provides an overview of measuring price and volume changes of the output of health and education providers. In the national accounts, outputs should reflect the results of production and these cannot normally be captured by outcome, the state of health or education of the population. However, we show that outcome information is required when it comes to quality adjustment of output measures. The paper clarifies terminology, and discusses output measurement and quality adjustment methods with a focus on health and education services.

**JEL Codes**: E01, I10, I20, O47

**Keywords**: health and education services, non-market output, output and outcome, quality adjustment

## 1. Introduction

A major purpose of the national accounts is to measure price and volume change of GDP and its major components as well as productivity. For this purpose, changes in volumes of output should be properly measured and reflect changes in quality. However, in national accounts practice, it occurs that input-based measures are used to track output of certain activities and no or insufficient quality adjustment is in place. A case in point is services such as education and health that are complex and for which there are often no economically significant prices. Often, methodologies have relied on measuring the volumes or the prices of inputs to obtain a measure of the volume or price of outputs. This means that it is impossible to capture productivity change. Ignoring productivity changes means foregoing important information for analysts and policy makers about two sectors that account for at least 10 percent of GDP in many OECD countries.

A recent review of OECD country practices of health and education output measurement (Schreyer, 2010) reveals a heterogeneous picture. In the field of education, about one third of all countries, mainly European as well as Australia and New Zealand, have already implemented output-based measures and others are planning to do so. In others, work is at the research stage. A similar picture emerges for health care although a larger number of countries have embraced output-based methods. That said, the categorization of a method as input- or output-based is not always clear cut and it is even more difficult to pass a judgment on the appropriateness of quality adjustment methods used. In sum, thus, a rather

*Correspondence to: Paul Schreyer, OECD Statistics Directorate, 2 rue André Pascal, 75116 Paris Cedex 16, France (Paul.Schreyer@oecd.org).


Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

disparate picture emerges by way of methods used, and consequently international comparisons have to be conducted with care.

In many countries, health and education services are provided by government or other non-market producers. It is clear that market production gives us direct observations on the value of production at current prices whereas the value of non-market production is typically estimated by summing up costs. But most of the tricky issues associated with the measurement of *volume* output apply to both market and non-market producers. In particular, the difficulties of keeping track of quality change and of entering and exiting products are present independently of the institutional affiliation of the producing unit. These problems are associated with the increasingly complex nature of modern services, for example in health care, and not with the question of whether these services are provided under market conditions or not. In consequence, many issues of output measurement are relevant for the activity as a whole, and not only under conditions of non-market provision, although non-market provision complicates measurement in several respects.

The discussion about the measurement of health and education services is by no means new. In 1975, Peter Hill developed principles and guidance for measuring health, education, and collective government services (Hill, 1975). More recently, the debate has resurfaced. Eurostat (2001) stated the desirability of applying output-based measures to non-market services. In the United Kingdom, the topic was taken up by the widely-discussed Atkinson Review (Atkinson, 2005). The measurement of services output and productivity has also been a longstanding topic of interest in the United States, with a series of publications including Triplett and Bosworth (2004) and Abraham and Mackie (2006). Health services in particular have been the subject of research on cost-effectiveness and productivity (Cutler *et al.*, 2006; Rosen and Cutler, 2007), just as education services have been scrutinized for efficiency and effectiveness. When analysis proceeds from a health care or educational policy perspective, measurement issues do not always surface, and where they do it is not always apparent how they relate to measurement principles in the national accounts. One objective of the present paper is to provide these links.[1]

The principle objective of our efforts here is to shed light on the measurement of volume measures of output for complex and economically important service activities, and we shall frame our discussion with examples from health and education, although they may apply more broadly. Terminology is important in this context and Section 2 starts by defining "output," "outcome," "inputs," and "activities." Sections 3–5 provide the economics of measurement in an increasingly complex setting. We move from the simple competitive case without quality

[1]One particular strand of productivity research is rooted in efficiency analysis where hospitals or educational establishments are compared using Data Envelopment Analysis and Stochastic Frontier Models as developed by Farrell (1957), Färe *et al.* (1985), and Aigner *et al.* (1977). These methods use either linear programming or econometric techniques to assess the relative productivity of establishments, with the added feature of distinguishing between efficiency effects (moves toward the efficiency frontier) and technical change (moves of the efficiency frontier). Also, DEA models do not necessarily require information on prices; quantity observations of inputs and outputs suffice. However, these inputs and outputs have to be well-defined to apply DEA and stochastic frontier models, and consequently they do not provide help in measuring inputs and outputs in the first place.

change (Section 3) to non-market production without and with quality change (Sections 4 and 5). Our main conclusion is that despite the fact that national accounts should focus on measuring outputs, in the presence of quality change, implicit or explicit information about outcomes is required. Section 5 also contains a proposal for the use of econometric techniques in the presence of non-market production. Section 6 concludes.

## 2. Terminology

Our point of reference is an economic unit that transforms volumes of inputs into volumes of outputs. *Inputs*, the goods and services to be transformed, comprise labor services, capital services, and intermediate inputs. Inputs are combined and transformed by way of a production technology. *Outputs* are suitably differentiated and are the number of constant-quality actions or activities (in the case of services), and the number of constant-quality physical units (in the case of goods).

*Outcome* is a state[2] that is valued by consumers—a functioning car, the state of health, the level of knowledge etc. Outcomes are influenced by many factors, and one of them may be the level of outputs. For example, the state of health (an outcome) is a function of medical care (output of the health industry), people's lifestyles, and the natural environment. Often, outcomes manifest themselves with a considerable lag to the provision of output as would be the case with long-term effects on human health. For all these reasons, outcomes are different from outputs. This difference between output and outcome has already been well described by Hill (1975): ". . . there is a danger that the purpose for which goods and services are to be used may be confused with the goods and services themselves. . . . Thus, in order to establish the nature of the output it is necessary to examine carefully precisely what good is actually exchanged between producer and consumer or what activity is actually performed by the producer unit for the benefit of the user unit" (p. 13). Outcome would then correspond to the purpose for which goods and services are to be used, and outputs to the goods and services themselves. In principle, the production boundary as defined in the national

[2]"Outcome" has been used in different ways in the relevant literature on non-market services. Two usages are common: in the health care literature, "outcome" is typically defined as the resulting change in health status that is directly attributable to the health care received. Triplett (1998) indicates this usage in the cost-effectiveness literature and quotes Gold *et al*. (1996), who define a health outcome as the end result of a medical intervention, or the change in health status associated with the intervention over some evaluation period or over the patient's lifetime. Employed in this sense, some authors suggest that the "output" of the health care industry should be measured by "outcome." Among national accountants, "outcome" is typically used to describe a state that consumers value, for example the health status without necessarily relating the change in this state to the medical intervention. Eurostat (2001) gives as examples of "outcome indicators" the level of education of the population, life expectancy, or the level of crime. Atkinson (2005) has the same usage of the word. Understood in this sense, outcome in itself cannot be a useful way to measure output or the effectiveness of the health or education system. In terms of national accounts semantics, the "marginal contribution of the health care industry to outcome" is the equivalent to the notion of "outcome" as used in the health care literature. As long as a particular definition is used consistently, the substance of the argument is of course unaffected and the only question is how useful a particular definition is for the purpose at hand. The present paper follows in the line of Eurostat (2001) and Atkinson (2005), and employs the term "outcome" in the sense of the national accounts literature.

accounts encompasses outputs, but not outcome. A superficial conclusion would therefore be that the national accounts statistician does not have to worry about outcomes, only outputs.

However, things are more complicated. While outcomes are different from outputs, they are not independent. One of the conclusions of the present paper will in fact be that it is virtually impossible—in particular for health and education services—to define quality-adjusted outputs without invoking outcomes one way or the other.

For the forthcoming discussion, it is useful to refine the definition of outputs and outcomes in two ways. First, outputs are broken down into two components: *activities or processes* and the *quality adjustment* applied to them. Processes are observable and countable actions by which services are delivered, although their characteristics may change over time. For education, a typical process measure is the number of pupils or the number of pupil hours taught in a particular grade. For health, a typical process measure is the number of treatments of a particular disease such as hip replacements. The purpose of quality adjustments applied to processes is to measure them in constant-quality units that are comparable over time or in space.

Note that the terminology above is different but consistent with terminology used in work relating to the health or education system *as a whole*: for instance, Rosen and Cutler (2007) develop a system of national health accounts that brings together inputs and outputs in the "production" of health. Medical care is only one input here; other inputs include the time people invest in health (such as exercise and sleep), or research and development. The output of such a health system is simply the state of health, "outcome" in our terminology. It is apparent, however, that the production of the health system as a whole is outside the production boundaries of the system of national accounts. Thus, while of significant analytical interest, it does not help the national accountant in his or her quest for better capturing output and productivity inside the established production boundaries.

## 3. Competitive Markets, No Quality Change in Products

Having dealt with terminology, we shall now turn to the theory of measurement. Our reference case here is a situation of competitive markets without quality change in products. We separately deal with the consumer and producer sides before bringing them together.

### 3.1. *The Consumer Side*

Our conceptual discussion starts with a simple market model of producers and consumers. For the moment, we take it that products are well-defined: output is expressed in observable, constant-quality counts of processes. On the demand side, consumers purchase the services supplied by producers. Standard economic theory attributes a utility function to consumers where utility (in unobserved "utils") depends on the quantity of goods and services consumed.

Let households' utility function be $U^t = U(H^t)$, where $U^t$ stands for the level of utility in period $t$ which depends positively on $H^t$, a state of the world that consumers value. $H^t$ corresponds to our notion of outcome. In general, $H^t$ will be vector-valued because there are many different states valued by consumers—for example, different aspects of the status of health, the level of knowledge, or the state of the natural environment. Consumers attach utility to a good or to a service because it affects *outcome*, i.e., a particular state that they value. We could also say that outcome is an intermediate step between consumption and utility and this is indeed the way it has been treated in the literature. In an application to health care, Berndt *et al.* (1998) distinguish between medical care ("output" in our terminology), the state of health ("outcome" in our terminology), and utility. They envisage a relationship whereby utility depends, among other variables, on the state of health, and where the state of health is itself dependent on health care services, on the environment, lifestyle etc.

If we follow this idea, outcomes depend, *inter alia*, on N different services consumed by households, and we shall label quantities of these services in period $t$ as $\mathbf{y^t} = [y_1^t, y_2^t, y_3^t, \ldots, y_N^t]$. However, outcome $H^t$ depends not only on services $\mathbf{y^t}$ that are purchased or obtained from producers but also on a host of other factors, $\Omega^t$. Examples include households' behavior with regard to smoking, alcohol consumption, or physical exercise. Or, in the case of education, $\Omega^t$ could stand for students' efforts or natural giftedness when $H^t$ stands for educational attainment and the state of knowledge of the population. For the moment, we formulate our "outcome" function and insert it into the utility function defined earlier:

(1)
$$H^t = H(\mathbf{y^t}, \Omega^t), \text{ and therefore}$$
$$U^t = U[H(\mathbf{y^t}, \Omega^t)] = V[\mathbf{y^t}, \Omega^t].$$

Households demand services to the point where prices equal the marginal utility generated by these services. In money terms, this gives $p_i^t = [\delta U / \delta y_i^t] / \lambda^t$ for $i = 1, \ldots, N$. Here, $1/\lambda^t$ is the marginal utility of income that is needed to convert "utils" into currency units. It is not difficult to see that the marginal contribution of a good or a service to utility is in fact a composite term, namely the marginal contribution of the product to outcome, $\delta H / \delta y_i^t$, multiplied by the marginal contribution of outcome to utility, $\delta U / \delta y_i^t = [\delta U / \delta H][\delta H / \delta y_i^t]$.

It follows that the marginal contribution to outcome for a particular product $i$, $\delta H / \delta y_i^t$, relative to the marginal contribution of another product $j$, equals the price ratio of the two products:

(2)
$$[\delta H / \delta y_i^t] / [\delta H / \delta y_j^t] = p_i^t / p_j^t.$$

The measurement implication of this relation is that there is a conceptual link between outcomes and *statistical classifications* for (consumer) goods and services. Given different products or product items (variants of a product), the question about their grouping or classification has to be answered with respect to the purpose of the price or volume index. The relation above suggests that from a consumer perspective, the criterion for grouping individual items is that they potentially offer the same contribution to outcomes, i.e., they satisfy the same or similar consumer needs. Put differently, they are substitutes from a consumer perspective. Conversely, if different items are not interchangeable from a consumer perspective, they should be treated as different products. In the presence of quality change or new and disappearing items, the question of grouping items becomes important. But the point to retain is that the organization of price or quantity measurement, in particular how products are classified and stratified, cannot proceed without some reference to outcome, if one wants to bring in a consumer perspective.

### 3.2. *The Producer Side*

We now turn to producers and take it that their production technology can be represented by a cost function[3] that shows the minimum costs required during a given period to produce a quantity of $N$ products $\mathbf{y^t} = [y_1^t, y_2^t, y_3^t, \ldots, y_N^t]$, for a given set of input prices and for a given technology. In the case of health service producers, a particular product could consist of a constant-quality treatment of a particular disease; in the case of education it could be the constant-quality hours of teaching provided for a particular grade. To keep the exposition tractable, it will be assumed that there is a technology set $S_i^t$ for each product that links output $y_i^t$ to a set of $M$ inputs, $\mathbf{x_i^t} = [x_{i1}^t, x_{i2}^t, \ldots, x_{iM}^t]$. Inputs are purchased at prices $\mathbf{w^t} = [w_1^t, w_2^t, \ldots, w_M^t]$ on factor markets where producers are price takers. A cost function for product $i$ can then be written as

$$(3) \qquad C_i^t = C_i^t(y_i, \mathbf{w}) = \min\{\mathbf{w} \cdot \mathbf{x} : (y_i, \mathbf{w}) \text{ is contained in } S_i^t\}; i = 1, \ldots, N.$$

A unit or average cost function for output type $i$ is defined as $c_i^t(y_i, \mathbf{w}) \equiv C_i^t(y_i, \mathbf{w})/y_i$ and measures the costs per unit of output $y_i$ during period $t$. With constant returns to scale which we shall assume for convenience here, unit costs are independent of the level of output, and average costs equal marginal costs. When producers minimize costs, it follows that total costs for a particular product are equal to minimum costs:

$$(4) \qquad C_i^t(y_i^t, \mathbf{w^t}) = c_i^t(\mathbf{w^t})y_i^t = \mathbf{w^t} \cdot \mathbf{x_i^t}.$$

We follow Diewert (2008) and define the producer's or sector's total cost function $C^t$ as the sum of cost functions for different products. This will be helpful in the definition of index numbers below:

$$(5) \qquad C^t(\mathbf{y^t}, \mathbf{w^t}) = \sum\nolimits_i^N C_i^t(y_i^t, \mathbf{w^t}) = \sum\nolimits_i^N c_i^t(\mathbf{w^t})y_i^t = \mathbf{w^t} \cdot \mathbf{x_i^t}.$$

Under constant returns to scale, and in a competitive market, producers provide services at the point where prices equal marginal costs. More generally, under market conditions, excess profits are competed away and prices will equal average costs so that:

$$(6) \qquad p_i^t = c_i^t(\mathbf{w_i^t}).$$

[3]For more information about cost functions see Shephard (1970) or Diewert (1974).

### 3.3. *Combining the Consumer and the Producer Side*

When the consumer and the producer side are combined, market equilibrium is characterized by

$$(7) \qquad [\delta U/\delta H][\delta H/\delta y_i^t]/\lambda^t = p_i^t = c_i^t(\mathbf{w^t}).$$

Thus, marginal utility from consumption and marginal costs of production are equal in equilibrium. There is of course nothing new about this statement. It is nonetheless useful because we see that in equilibrium, consumer valuation and producer valuation of a good or of a service coincide at the margin when there is a market-clearing price. The implication is that when weights are needed to aggregate across products, there is no need to invoke either a consumer or a producer perspective—the value of market transactions is all that is needed and it combines the two sides of the market.[4] Also, in the absence of quality change and as long as there are no new or exiting products, it would appear that there is little to worry about the distinction between output and outcome—market prices jointly value the marginal contribution of consumption to outcome and also reflect marginal production costs.

## 4. Non-Market Production, No Quality Change

Having established the competitive case with a well-defined set of products that does not change over time, we shall now direct attention to the case of non-market production. When goods or services are provided by non-market producers, they are provided at a price that does not cover costs and which may even be zero. In this case, the price at which products are transacted loses its significance as an indicator of marginal or average costs. Nor is the price necessarily linked to a utility-maximizing quantity of consumer demand, as was the case in the competitive environment. Therefore, our convenient link between the producer and the consumer side, established in the first section, breaks down:

$$(8) \qquad [\delta U/\delta y_i^t]/\lambda^t \neq c_i^t(\mathbf{w^t}).$$

An immediate consequence from this situation is that the well-established body of literature on the theory of producer price indices, notably Fisher and Shell (1972), Archibald (1977), and International Monetary Fund *et al.* (2004) no longer applies. The theory of the output price index relies on revenue functions for

[4]This is a simplification. In practice, and in the presence of transport costs or taxes, there is no unique market price and a distinction has to be made between different valuations. For example, from a consumer perspective, a valuation at purchasers' prices is appropriate, which is inclusive of taxes and transportation margins. From a producer perspective, a valuation at basic prices would be more appropriate, which excludes, for example, taxes payable and subsidies receivable in conjunction with production or sale. The statement in the text is also a simplification in the sense that output price indices for producers and input price indices for consumers are typically compiled at different levels of aggregation and on the basis of different classifications. This may also lead to differences in weights between output price indices for producers and input price indices for consumers. However, even when aggregation happens differently in the producer and in the consumer case, the price at the lowest level of aggregation at which the transaction takes place is a market price and reflects the joint influences of producers and consumers.

producers and stipulates revenue-maximizing behavior, given a set of market prices for producers' outputs. In a non-market environment,[5] this is not a useful assumption, and consequently, revenue functions cannot be used as a conceptual basis for output price indices.

However, measurement can be based on *unit costs or quasi prices*. When transaction prices are significantly below cost or zero, it is customary in the national accounts to measure the money value of output as the sum of costs. One could also say that output is valued at quasi prices. They are those (unobserved) prices that emulate a competitive situation where prices equal average costs per product. With unit costs at hand, they can be treated *as if they were prices*:

$$p_i^t \equiv c_i^t(\mathbf{w^t}). \tag{9}$$

If we maintain the (courageous) assumption[6] that non-market producers are cost-minimizing units, then minimum costs equal actual costs or $c_i^t(\mathbf{w^t})y_i^t = \mathbf{w^t} \cdot \mathbf{x_i^t}$, and it follows that

$$p_i^t y_i^t \equiv c_i^t(\mathbf{w^t}) \cdot y_i^t = \mathbf{w^t} \cdot \mathbf{x_i^t}. \tag{10}$$

Expression (10) states the obvious, namely that with quasi prices, the value output of product *i* equals the value of inputs used in production of product *i*. This is the way non-market output is valued in the *System of National Accounts*.[7] What is important for the purpose at hand is the fact that this equality of inputs and outputs in value does *not* imply equality of inputs and outputs in volume or quantity. If this were the case, our efforts to derive volume measures of output that are separate from volume measures of inputs would be put in question and with it any attempt to measure productivity change in non-market production.

The main difference between cost-based prices of outputs and prices of inputs is that the former correspond to *costs per unit of output* (such as the costs for one treatment of a heart attack or the costs for one year of schooling), whereas the latter correspond to the *costs per unit of input* (such as wages per hour of a nurse or the salary of a teacher).

[5]Note that despite the fact that our discussion has been couched in terms of non-market producers, it carries over to the more general case of regulated industries. For example, Lawrence and Diewert (2006) measure the quantity index of output for New Zealand electricity utilities with a cost-based index because there are no meaningful revenue shares or prices for the three types of outputs identified for utilities: throughput of electricity, system line capacity, and connections.

[6]The main advantage of this assumption is that it allows a conceptually clean identification of productivity change with technical change. Suppose that cost changes are broken down into changes in the quantity of output, changes in the prices of inputs, and productivity change as in Diewert (2008). Then, one of the conditions that the measured productivity change is a reflection of technical change (or a shift of the cost function) is that the assumption of cost-minimizing producers holds. If this assumption is relaxed, the measured productivity change can encompass several elements, in particular a movement toward or away from the efficiency frontier and a movement of the cost function itself. As our present concern is with the measurement of output and only indirectly with the measurement of productivity, we have maintained the assumption of cost-minimization for reasons of simplicity. Giving up this assumption would complicate the exposition and would require more discussion by way of applicable index number formulae. See also Balk (1998) for productivity measurement when producers are not acting as cost minimizers.

[7]For a genesis of the treatment of non-market production in the national accounts and the many issues associated with it, see Vanoli (2002).

Diewert (2008) shows formally how a cost-based volume index of output can be defined. He proposes a family of cost-based output quantity indices and focuses on the Laspeyres ($Q_L$), on the Paasche ($Q_P$), and on the Fisher ($Q_F$) case. In line with the economic approach toward index numbers, Diewert defines the Laspeyres version of a cost-based output quantity index as the (hypothetical) total cost $C^0(\mathbf{y}^1, \mathbf{w}^0)$ of producing the output vector $\mathbf{y}^1$ of period 1 under the conditions of period 0 technology and input prices, divided by the actual costs of period 0, $C^0(\mathbf{y}^0, \mathbf{w}^0)$. Similarly, he defines a Paasche type index as the actual costs of period 1, $C^1(\mathbf{y}^1, \mathbf{w}^1)$, divided by the hypothetical costs $C^1(\mathbf{y}^0, \mathbf{w}^1)$ that would have been incurred, had the products of period 0 been produced in period 1, under the technological constraints of period 1 and given period 1 input prices:

$$
\begin{aligned}
(11)\qquad Q_L &= C^0(\mathbf{y}^1,\mathbf{w}^0)/C^0(\mathbf{y}^0,\mathbf{w}^0) = \sum\nolimits_i^N c_i^0 y_i^1 \Big/ \sum\nolimits_i^N c_i^0 y_i^0 \\
Q_P &= C^1(\mathbf{y}^1,\mathbf{w}^1)/C^1(\mathbf{y}^0,\mathbf{w}^1) = \sum\nolimits_i^N c_i^1 y_i^1 \Big/ \sum\nolimits_i^N c_i^1 y_i^0 \\
Q_F &= [Q_L Q_P]^{1/2}.
\end{aligned}
$$

For a number of practical reasons, we prefer working with a "price" rather than a quantity index *à la* Diewert for non-market producers and then deflate total costs by a price index. To this end, we construct an *indirect index of quasi prices* by dividing total costs by the volume index of output:

$$
\begin{aligned}
(12)\qquad P_L &= [C^1(\mathbf{y}^1,\mathbf{w}^1)/C^0(\mathbf{y}^0,\mathbf{w}^0)]/Q_P = \sum\nolimits_i^N c_i^1 y_i^0 \Big/ \sum\nolimits_i^N c_i^0 y_i^0 \\
P_P &= [C^1(\mathbf{y}^1,\mathbf{w}^1)/C^0(\mathbf{y}^0,\mathbf{w}^0)]/Q_L = \sum\nolimits_i^N c_i^1 y_i^1 \Big/ \sum\nolimits_i^N c_i^0 y_i^1 \\
P_F &= [P_L P_P]^{1/2}.
\end{aligned}
$$

A useful interpretation of this quasi-price index can be obtained by rewriting the Laspeyres or Paasche version in expression (12). For example, after inserting the theoretical expression for $Q_P$ into the first line of (12), $P_L$ can be presented as the product of two terms:

$$
\begin{aligned}
(13)\qquad P_L &= [C^1(\mathbf{y}^1,\mathbf{w}^1)/C^0(\mathbf{y}^0,\mathbf{w}^0)]/Q_P \\
&= [C^1(\mathbf{y}^1,\mathbf{w}^1)/C^0(\mathbf{y}^0,\mathbf{w}^0)]/[C^1(\mathbf{y}^1,\mathbf{w}^1)/C^1(\mathbf{y}^0,\mathbf{w}^1)] \\
&= [C^1(\mathbf{y}^0,\mathbf{w}^1)/C^0(\mathbf{y}^0,\mathbf{w}^0)] \\
&= [C^1(\mathbf{y}^0,\mathbf{w}^1)/C^1(\mathbf{y}^0,\mathbf{w}^0)][C^1(\mathbf{y}^0,\mathbf{w}^0)/C^0(\mathbf{y}^0,\mathbf{w}^0)].
\end{aligned}
$$

The first term in the last line of (13) is an economic index of input prices: costs are compared between two situations, with technology and the level of output held fixed, but input prices allowed to vary. The second term in the same line is an inverted productivity index: for a given reference output and input prices, changes in minimum costs between the periods are compared. Similar transformations could be applied to $P_P$ and then combined with $P_L$ to yield a decomposition of $P_F$, but there is no need to present them here. The main point can easily be explained with the decomposition of $P_L$ only: in a market situation, a productivity index equals an input price index divided by an (output) price index: if output prices rise less rapidly than input prices, this implies productivity improvements. In the present non-market case, the quasi-price index for outputs plays a similar role as the output price index in a market situation. If quasi prices (unit costs) rise less rapidly than input prices, there has been productivity change.

The measurement of productivity as a shift in the cost function is a well-established methodology. Whenever there are situations of imperfect competition, or non-constant returns to scale in production, there are at least two ways of measuring productivity change—as a shift of the production possibility curve or a shift of the cost function.[8] The two approaches do not generally yield the same result and there is no strong *a priori* reason to prefer one over the other. We conclude that the cost-based productivity measure is a fully valid measure of technical change for non-market producers.

But (13) also shows that despite the fact that much of the discussion about non-market producers has been by way of costs, we *are* lending an output perspective to our calculations: unit costs or quasi prices are productivity-adjusted input prices and the productivity adjustment marks the movement from an input perspective toward an output perspective in measuring non-market activity. This is not always well understood, because costs are rightly seen as input-related variables. But the above makes it clear that considering costs per unit *of output* differentiates an output perspective from considering costs per unit *of input*, i.e., the input perspective. The output perspective remains a proxy only insofar as it brings in no direct consumer valuation—unit costs are not a product of the interplay between producers and consumers as in the market case. Unit costs are only reflective of the supply side. However, some elements of consumer valuation enter the picture when it comes to quality adjustment, and this is the topic of the following section.

Recent work by the OECD (Koechlin *et al.*, 2010) follows this avenue. Quasi prices for hospital services were collected across a sample of OECD countries. With cost weights attached to each type of service, these price comparisons constitute purchasing power parities for hospital services that can be used as a spatial deflator of national nominal expenditure on hospital services. Such output-based estimates help to answer the question of whether differences in health expenditure between countries reflect differences in the volume of health care provision, or differences in the quasi price or unit cost of treatments.

## 5. Non-Market Production and Quality Change

### 5.1. *Outcome, Stratification, and Implicit Quality Adjustment*

There is an extensive literature on how to deal with quality change in existing products, with the exit of old products and with the entry of new products if one wants to compile price or quantity indices. Quality change counts among the most challenging measurement faced by price statisticians. Early references include Stone (1956) and Griliches (1971), and more recent ones International Monetary Fund *et al.* (2004) and Triplett (2006). The reader is referred to these volumes for

[8]Balk (1998) provides a full treatment of the various productivity measures. In his terminology, our measure of technical change would be labeled a "dual input based technical change index" (p. 58). Diewert and Nakamura (2007) also discuss dual, cost based measures of productivity change.

a more complete treatment of quality adjustment methods. The task in this section is twofold: to discuss how the measurement of quality change relates to outputs and outcomes; and to propose a method for quality adjustment for the non-market case. To start, we cite Triplett (2006, p. 15) and recall some key principles and methods that are followed in measuring quality change.

> Agencies that estimate price indexes employ, near universally, one fundamental methodological principle. The agency chooses a sample of sellers . . . and of products. It collects a price in the initial period for each of the products selected. Then, at some second period, it collects the price for exactly the same product, from the same seller, that was selected in the initial period. The price index is computed by matching the price for the second period with the initial price, observation by observation or "model by model" as it is often somewhat inaccurately called.

One technique to deal with quality change in products is thus to group them such that only products of the same specification are compared over time or in space. Such grouping or matching ensures that only prices or quantities of products of the same or very similar quality are compared. The idea is that products of different quality are treated as different products. Examples for such grouping in education are establishments that provide different services, such as boarding schools as opposed to day-time schools or hospitals with different levels of non-medical services. The (quasi) price of a particular medical treatment would then be followed inside the group of establishments of a particular type. This is a way of controlling for quality characteristics and to track a price for a constant quality service.

Note, however, that grouping also relies on an important assumption: to show a price or quantity movement that is representative of a product group, the price or quantity movements of those products that *are* matched have to be a good indicator of the price or quantity movements of those products that are *not* matched—in particular products that are newly entering the market. Price or quantity changes that arise outside the sample of matched products are ignored.

The non-market case on which our attention is focused here shows also the importance of choosing the right level of aggregation where matching takes place. And it is again considerations of outcome that govern this choice. Take the following case. Two medical procedures are considered, of different unit costs, but equally interesting from the consumer's viewpoint—both procedures treat the same medical problem equally well. In other words, the contribution of each treatment to outcome, from a consumer perspective, is the same. Treating each procedure as a separate product, i.e., setting the elementary level for the construction of the price or volume index below each procedure, can lead to counter-intuitive results: more of the cheaper but equally helpful treatment translates into a decline in the volume of output when, as would often be the case, the new procedure only gradually replaces the old procedure. This is best illustrated by way of an example (Schreyer, 2010). Suppose there are two treatments for a disease, traditional surgery and laser treatment, and assume that laser treatment is introduced in period 1 (see Table 1). In addition, as may well be the case, the unit cost

TABLE 1
VOLUME INDICES AND QUALITY CHANGE; A NUMERICAL EXAMPLE

| | Traditional Surgery | | | Laser Surgery | | |
|---|---|---|---|---|---|---|
| | Period 0 | Period 1 | Period 2 | Period 0 | Period 1 | Period 2 |
| Unit cost | 100 | 100 | 100 | – | 90 | 90 |
| Number of interventions | 50 | 40 | 5 | 0 | 10 | 45 |
| Total cost | 5000 | 4000 | 500 | 0 | 900 | 4050 |

of laser treatment is lower than the unit cost of traditional surgery. The total number of interventions in each period remains the same.

Now consider a (simplified) matched-model approach toward calculating a volume change from period 0 to period 1. In the simplest case, the volume index is given by the quantity changes in the two treatments, each weighted by the cost share it occupies in period 0. As laser surgery does not yet exist in period 0, it receives a zero weight so the volume index of treatments is simply the change in the number of traditional surgery interventions, or $(40/50 - 1) = -20\%$. Between periods 1 and 2, the corresponding volume index equals $[s_T(5/40) + s_L(45/10)] - 1 = -7.1\%$, where $s_T = 82\%$ and $s_L = 18\%$ are the period 1 cost shares of the traditional and laser treatments, respectively. This approach treats the two treatments as different products, and the sharp drop in the total volume index in period 1 reflects the "new goods" problem that arises when new products enter the sample that cannot be compared with quantities in the base period. The implicit assumption in this model is that consumer valuation of the two products is captured by the relative unit costs, so if laser surgery is cheaper than traditional surgery, the method implicitly quality-adjusts downward the quantity of laser surgery when combined with traditional surgery. In a perfect market, the price of the traditional treatment would see an instantaneous downward adjustment, bringing consumer valuation of the two processes in line. A different result arises when it is considered that the two treatments are perfect substitutes, i.e., that they are in fact the same product. In this case, no cost weighting is applied between the two treatments—and the number of treatments is simply added up. As there are 50 interventions in every period, the result is a volume index that shows zero growth and a declining price index, reflecting the drop in average unit costs of treatment. The previous method is justified if consumers are indifferent to the two treatments. If this is not the case, and they prefer laser over traditional surgery because the former is less intrusive or requires fewer days of recovery, an explicit quality-adjustment is needed. Such an adjustment can be applied to the quantity measures, either by scaling up the quantity of laser treatments or by scaling down the quantity of traditional treatments. Whichever way this is done, the implication is always that one treatment is expressed in equivalents of the other, and the ratio should in some way reflect consumer preferences. Alternatively, prices or unit costs could be rescaled before constructing a price index. Suppose the adjustment factor is 1.1—each laser treatment is the equivalent of 1.1 traditional treatments. Then, expressed in "traditional surgery-equivalents," the number of treatments is 50 in period 0, $40 + 10 \cdot 1.1 = 51$ in period 1 and $5 + 45 \cdot 1.1 = 54.5$ in period 2. The resulting volume index is +2% in period 1 and +6.9% in period 2. Obviously, the difficulty lies in determining the adjustment factor which should: (i) reflect consumer preferences; and (ii) be unidimensional.

We conclude that even in a situation where matching between products is perfect, and no explicit quality adjustment may be needed, *some* reasoning about outcome is in place, if only to group substitutable products together in one stratum. No judgment about substitutability can be made without invoking, at least implicitly, some judgment about outcomes.

But matching is rarely perfect. Matching of the quasi prices or volumes of non-market services such as health and education services is unlikely to control completely for particular characteristics associated with the provision of these services. Or a representative service may change its characteristics, akin to a product that price collectors are no more able to find in a particular outlet and that has to be replaced with a new product. These are the instances where explicit quality adjustment comes into play, of which more later.

### 5.2. *Outcome and Explicit Quality Adjustment*

If matching is insufficient to control for key characteristics of service provision, other, explicit, techniques have to be invoked to account for quality change. In general, the quality of a product can be expressed by the quantity of its characteristics. Quality change can then be captured by the change in characteristics. Similarly, price changes in products can be attributed to pure price changes and to those price changes that reflect changes in product characteristics. This is the approach followed by hedonic price indices[9] that are now well established among statistical agencies.

Quality adjustments require the identification of a set of characteristics such as the speed, engine size, or equipment for a car, or the processor speed for a computer. Cutler and Berndt (2001) use patient characteristics, information on different types of depression, variables on medication and the like to estimate a hedonic price model for the treatment of depression; the idea being to isolate those price changes that are due to changes in characteristics from those price changes that constitute "inflation." An important result of the hedonic model is that it allows the identification of characteristics and provides a market valuation of each one.[10] Market valuation, in turn, is a convenient way of aggregating across characteristics because everything is expressed in a single monetary unit.

At a first glance, hedonic regressions would appear to be badly suited for a non-market environment, given that hedonic methods are all about extracting "market" information from examining the link between market prices and product characteristics. At a second glance, however, the basic principles of hedonic regressions *would* appear to be adaptable for a non-market environment, albeit with a different interpretation.

[9]See Triplett (2006) for a comprehensive discussion and a full set of references.

[10]Rosen (1974) demonstrated that in general, those characteristics of a product will show up in the function that are valued by consumers *and* that have cost implications for producers.

To demonstrate this idea, we shall augment the simple producer model above by product characteristics. Thus, for every product $y_i^t (i = 1, 2, \ldots, N)$ there is a vector of $n(i)$ characteristics $\mathbf{z_i^t} \equiv [z_{i1}^t, z_{i2}^t, \ldots, z_{in(i)}^t]$ that qualifies this product. Characteristics play a double role.

First, they help dealing with heterogeneity. Truly homogenous products are rare if they exist at all, and every type of product contains a more or less important element of heterogeneity. For example, if health product $y_1$ is "treatment of a heart attack," this will encompass strokes of different severity, and patients of different age suffering from a stroke. If old patients require more intense treatment than young patients, or if more severe strokes necessitate more intense care than less severe strokes, there are in fact different services involved. To some degree, this can be accommodated by stratification and matching (see above), but only up to a point. Additional heterogeneity is best captured by identifying, through knowledge of the product, those characteristics that make one service distinct from another. The variable $z_{11}$ could thus be "age of patient," the variable $z_{12}$ "degree of severity of stroke," and so on.

Second, characteristics help dealing with quality differences. In the case of health services, for example, observations on the same type of treatment may come from different institutions, for which there exist additional explicit information on characteristics that are considered attributes of quality from a consumer perspective, for example the in-hospital waiting time after a hip fracture.[11] Waiting time is an example of a direct measure of a quality characteristic. Other characteristics may only be captured indirectly. For example, the treatment of a particular disease may or may not include the use of a scanner. If there are reasons to believe that the use of certain types of modern equipment increases the quality of diagnosis, this is a relevant variable to be picked up. Yet another type of characteristic may have to do with secondary products or amenities provided by the unit that produces services. For example, the same medical service can be provided in hospitals with different types of accommodation services: multiple versus single-patient rooms, quality of meals, equipment of rooms with phones and TV sets, etc. The total average costs per treatment in two different establishments cannot be compared unless one controls for these characteristics. A similar case can be made for education establishments with regard to the provision of meals in school or the leisure activities on offer.

An important characteristic for the health and education area is the number and quality of personnel per patient or per student. Class size, for example, has long been quoted as an important quality characteristic of schools. Waiting time, if not measured directly, can probably be tracked well by the number of nurses or doctors per patient, and is an obvious quality characteristic for medical services. However, we have to be careful in introducing input measures lest there be a danger of falling back into measuring the volume of outputs by the volume of inputs. And this is exactly what we had set out to get away from in the first place! Which characteristics are admissible? And how should they be introduced?

In a market situation, the relevant characteristics are those that are valued by consumers and which also have a cost associated with their provision. In the first

[11]This is an example from the OECD Health Care Quality Indicators (see Armesto *et al.*, 2007).

place, this means they are valued by consumers but it also implies that there is a cost associated with the provision of characteristics, otherwise they would have no impact on prices. By analogy then, in the non-market case, the choice of characteristics should be such that they are primarily relevant for consumers, i.e., they are utility-enhancing. Characteristics that are only or predominantly relevant for producers, such as whether catering is carried out in-house or whether it is outsourced, should not be candidates for characteristics. There is no single, universal rule here and the choice of characteristics, as in all hedonic regression studies, is a matter of knowledge of the product. The choice of characteristics brings us back to the notion of outcome. Identification of relevant characteristics cannot proceed without recourse to outcome—only outcome considerations from a consumer perspective will help getting on with the choice of quality characteristics to be included in the hedonic regression in a non-market context.

It is worth noting that this is not an issue specific to non-market services, although it may be more accentuated in this context. Empirical difficulties with the choice of characteristics are also an important issue for market-provided products, such as computers in Triplett (2006, p. 167):

> In the end, economic theory does not specify the characteristics. Choosing the characteristics requires marketing or engineering or other information about the product and what buyers want to do with it.

The difference to the non-market case is that in a market context, the hedonic regression provides a market valuation of the characteristic in question, i.e., a valuation that arises as the intersection between consumers and producers. In the non-market context, the valuation is confined to the cost side. It could be seen as the marginal valuation of the characteristic by the non-market institution or by the government through its choice of allocating resources to different characteristics.

Having discussed the nature of characteristics $\mathbf{z}_\mathbf{i}^\mathbf{t}$, we shall now link them to the cost function that constitutes the measurement framework. Recall that for every product $i$, there was a unit cost function $c_i^t(\mathbf{w}^t)$. $y_i^t$ remains the right measure for the volume output but we now have to allow for the fact that $y_i^t$ may be not directly observable. We shall call the observable, but not quality-adjusted measure of output, $\zeta_i^t$. For example, $\zeta_i^t$ could be the total number of stroke treatments in different hospitals, irrespective of the age of patients or of the severity of cases treated. To model quality change one can say that the quality-adjusted, but unobserved output $y_i^t$ equals a measure of unadjusted but observable output $\zeta_i^t$, multiplied by a quality-adjustment function $g_i^t$, which in turn depends on the characteristics $\mathbf{z}_\mathbf{i}^\mathbf{t}$ that attach to each product. Characteristics should be defined in such a way that a bigger quantity of characteristics is associated with higher quality.

(14) $$y_i^t = \zeta_i^t g_i^t(\mathbf{z}_\mathbf{i}^\mathbf{t}).$$

Total costs for a non-market product should remain invariant to how they are split up into a unit cost and a quantity component. This requires that

(15) $$c_i^t(\mathbf{w}^\mathbf{t}) y_i^t = u_i^t \zeta_i^t,$$

where $u_i^t \equiv c_i^t y_i^t / \zeta_i^t$ has been defined as the unadjusted unit cost for product *i*. In combination with (14) one obtains

$$(16)\qquad \begin{aligned} u_i^t &= c_i^t(\mathbf{w^t}) y_i^t / \zeta_i^t \\ &= c_i^t(\mathbf{w^t}) g_i^t(\mathbf{z_i^t}), \text{ or, taking logs,} \\ \ln u_i^t &= \ln c_i^t(\mathbf{w^t}) + \ln g_i^t(\mathbf{z_i^t}). \end{aligned}$$

The last expression resembles a hedonic function.[12] Quality-unadjusted measures of unit costs of non-market output are broken down into a term that reflects the true, quality-adjusted unit costs and a term that captures the effects of characteristics. When set up as a hedonic regression, the relationship (16) constitutes an empirical avenue toward correcting for differences in quality or more generally, for differences in characteristics between different observations within the same type of product. There is a vast literature on hedonic regressions and price measurement,[13] and we shall not expand on it here beyond providing a sketch for an implementation method.

In a non-market context, one can think about $\{u_{ik}^t\}$ as a set of observations on unadjusted unit costs for a particular product, for example costs per cataract treatment in periods $t = 0, 1$ from a cross-section of hospitals, where each observation is indexed by $k = 1, 2, \ldots, K_i$. A hedonic regression that uses the dummy variable method then takes the following form:

$$(17)\qquad \ln u_{ik}^t = \beta_0 + \alpha D^t + \sum\nolimits_m^{n(i)} \beta_{im} \ln z_{imk}^t + \varepsilon_k^t.$$

Here, the hedonic equation in the last line of (16) has been transformed into an implementable form by setting the quality-adjusted unit costs $\ln c_i^t$ to equal the regression parameters $\beta_0 + \alpha D^t$, where $D^t$ is a dummy variable that equals 1 in period $t = 1$and that equals zero in period $t = 0$. The quality adjustment term ln $g_i^t(\mathbf{z_i^t})$ has been specified as $\sum_m^{n(i)} \beta_{im} \ln z_{imk}^t$ and a statistical error term $\varepsilon_k^t$ has been added. With suitable econometric methods, parameter estimates of the various coefficients are obtained. Of particular interest here is the parameter $\alpha$ that comes with the time dummy[14] variable. In antilog form, it represents the quality-adjusted index[15] of unit costs for product *i*:

$$(18)\qquad \ln c_i^1 - \ln c_i^0 = (\beta_0 + \alpha) - \beta_0 = \alpha.$$

[12]"A hedonic function is a relation between prices of varieties or models of heterogenous goods or services and the quantities of characteristics contained in them" (Triplett, 2006, p. 229).

[13]Court (1939) is frequently credited to have been one of the first researchers to use a hedonic function for purposes of price measurement.

[14]To simplify notation, we made no distinction between theoretical parameter values $\beta_0$, $\alpha$ etc. and estimated parameter values as they are obtained from the regression.

[15]Triplett and McDonald (1977) showed that the specification of the hedonic regression implies a certain index number formula in the aggregation of the individual observations. For the double-log specification used here, one gets $\exp(\alpha) = [\Pi_k (u_{ik}^1)^{1/K(1)} / \Pi_k (u_{ik}^0)^{1/K(0)}] / \text{QA}$. QA is a quality adjustment term of the following form $\text{QA} = \exp\{\Sigma_n \beta_{in} [\Sigma_k \ln z_{ink}^1 / K^1 - \Sigma_k \ln z_{ink}^0 / K^0]\}$. This is a term that captures differences in characteristics between the two periods, where each difference gets weighted with the corresponding regression coefficient. In the absence of changes in characteristics between the two periods, and no change in the number of observations, QA equals unity.

Hedonic coefficients are thus a tool for estimating the quality-adjusted rate of change in unit costs. In a market context, hedonic coefficients reflect the valuation of characteristics by the interaction of consumers and producers as shown by Rosen (1974). When competition is imperfect, interpreting coefficients becomes more difficult, because in addition to consumer preferences and producer costs, the degree of market power affects the size and possibly the sign of coefficients, as discussed by Pakes (2003). In a non-market setting, there is no direct market interaction between consumers and producers. Resources are allocated by producers, more or less in response to demand, but the valuation of characteristics is a cost valuation. Thus, coefficients in the hedonic equation above provide an indication about the technology of supply of characteristics and its cost structure. Supply in a non-market context does not directly interact with demand. In such a situation, regression coefficients are indicative of costs or producer valuation. However, despite the fact that the *valuation* of characteristics is a cost valuation, the *choice* of characteristics should largely reflect a consumer perspective as was explained earlier.

The above sketch of a hedonic method has not been tested and conveniently ignores the many empirical problems that its implementation would likely have to face. Data requirements are probably enormous and a core question, the choice of characteristics, makes this approach a challenging task. Nonetheless, exploratory studies would be helpful to assess feasibility and usefulness.

### 5.3. *Output as the Marginal Contribution to Outcome*

One consequence of (8) was the inapplicability of the theory of output price indices. We solved this by reverting to unit costs or quasi prices. In so doing, we implicitly signed up to another decision on how to deal with the inequality of marginal utility and marginal costs: quasi prices or cost-based output volumes *à la* Diewert (2008) imply that weights are cost shares of goods and services, and not expenditure shares that reflect consumer utility. When it came to quality change, we stuck with a cost-based approach but applied explicit quality adjustment to unit costs, and implicitly, to volume measures of output. Outcomes played an indirect role only in the sense that they helped stratifying activity counts and choosing the set of characteristics for quality adjustment.

An alternative way of constructing volume and price indices of output consists of directly invoking the consumer perspective and constructing measures of output that are based on a product's marginal contribution to outcome. Atkinson (2005) puts this forward as one of the methods of measuring the output of non-market producers. In our notation, this would mean tracking movements of output through observation of movements of outcomes caused by the change in output, i.e., $\delta H/\delta y_i^t$. One obvious advantage of this procedure is that—at least in concept—it provides a solution to the problem of quality adjustment. If all that consumers care about is outcome, and we are able to trace outcome and measure its change, then there is no need to explicitly quality-adjust measures of production.

Under an outcome-based approach, the volume of output is taken as the marginal contribution of the product to outcome as perceived by the consumer. Consider first a non-market producer, whose value of production has been set to equal costs. With this constraint, total costs would equal the total value of each product's contribution to outcome:

$$\text{(19)} \qquad \text{Total value of production}_i^t = u_i^t \zeta_i^t = [(\delta U/\delta H^t)/\lambda^t][(\delta H^t/\delta y_i^t)y_i^t].$$

As before, we allow for quality change in the provision of goods or services. Thus, a direct comparison of unadjusted unit costs over time, $u_i^1/u_i^0$, would give rise to a biased price index, and a comparison of unadjusted quantities, $\zeta_i^1/\zeta_i^0$, would give rise to a biased volume index because the quality of the observed units $\zeta_i^t$ changes over time. At the same time, all that can be directly observed are the unadjusted unit costs along with the unadjusted quantities. An outcome-based approach toward measuring prices and volumes deals with quality change in the following way: the money value of the contribution to utility that a change in outcome generates, is taken as the "true" price $p_W^t \equiv (\delta U/\delta H^t)/\lambda^t$. On the volume side, the marginal contribution of the good or service to outcome would constitute the "true" quantity $y_{W,i}^t \equiv (\delta H^t/\delta y_i^t)y_i^t$. By way of example from health services, $p_W^t$ would correspond to the net gains in welfare from an additional life-year and $y_{W,i}^t$ would correspond to the number of life-years gained from a medical service of type *i*. Note that when the usual assumption of declining marginal utility applies, the true price of health services $p_W^t$ declines with an improved state of health, as $\delta^2 U/\delta H^2 < 0$, everything else being equal. Although not directly observable from transactions on the market, these "true" prices and quantities can in principle be estimated.

In the area of education, the seminal work by Jorgenson and Fraumeni (1989) is an example where the marginal contribution to outcome in the form of additions to human capital due to education is taken as the output measure of the education sector. In the area of health care, there is an increasing number of studies that have used measures of outcome to value health care output.[16] For example, Cutler *et al.* (1998) pursue this research avenue and derive a price index for heart attack treatment. Conceptually, their cost-of-living index is comparable to the change in $p_W^t$. It is measured as (one minus) the net benefit from a change in heart attack treatment. If the benefits from medical practice changes are greater than the costs, consumers are better off and the cost of living declines. Empirically, then, the authors measure $p_W^1/p_W^0 - 1$ via estimating the net benefits from improvement in health due to the treatment of heart attacks. Given an estimate for $p_W^1/p_W^0$ and given the constraint (11), we can work out the implied quality adjustment factor to the observed quantities:

$$\text{(20)} \qquad y_{W,i}^1/y_{W,i}^0 = [\zeta_i^1/\zeta_i^0][(u_i^1/u_i^0)/(p_W^1/p_W^0)].$$

This shows that the index of quality-adjusted volumes of medical services $y_{W,i}^1/y_{W,i}^0$ corresponds to the index of unadjusted quantities $\zeta_i^1/\zeta_i^0$ times a quality

[16]See also the volume edited by Cutler and Berndt (2001) for other examples of new medical care price indices.

adjustment factor $(u_i^1/u_i^0)/(p_W^1/p_W^0)$ that is just the ratio between the index of unadjusted unit costs and the cost of living index.[17]

It is useful to compare this result with the hedonic adjustment described in the previous section. Expressions (16) and (18) taken together yield a formally similar quality adjustment where the regression coefficient $\alpha$ replaces the cost of living index $p_W^1/p_W^0$:

$$(21) \qquad y_{,i}^1/y_{,i}^0 = [\zeta_i^1/\zeta_i^0][(u_i^1/u_i^0)/\exp(\alpha)].$$

Despite the formal similarity, there is of course no reason why the two adjustments should be empirically equivalent. The quality adjustment factor in (20) is based on a non-parametric calculation that relies on a one-dimensional measure—outcome—that is supposed to subsume all quality aspects of production. The quality adjustment in (21), on the other hand, is based on a parametric calculation that uses multi-dimensional quality characteristics, to which cost weights are applied. The link to outcome remains indirect in that only the choice of characteristics reflects considerations of outcome, but no direct measure of outcome enters the set of characteristics.

There are many practical considerations before any of these approaches can be implemented, and output measures that rely on the marginal contribution to outcome such as in (20) are a long way from being regular features in statistical programs. Note the following empirical issues.

It is not obvious how to identify the marginal impact of output to outcome. As was mentioned earlier in this paper, outcomes (for example, the state of health) are affected by many factors, not only by the good or service under consideration. Thus, the effect of health care on the state of health should not be affected by any other factors that influence consumer outcome, such as the lifestyle of patients. It is empirically difficult to control for these other influences on outcome. For example, Cutler *et al.* (2006), in a major study on the value of medical spending in the United States, have to rely on common sense arguments and some general results from medical research to estimate the part of improved life expectancy that is attributable to medical care.[18]

A more subtle issue is the following: a measure of the contribution to outcomes should reflect the normal, or expected effect of the activity whose output is to be measured. "Normal," "average," or "expected" effects should be considered rather than ex-post, individual effects. This distinction arises in the context of service provision where the consumer is typically actively involved in the provision

[17]Triplett (1998), referring to the cost-of-living adjustment by Cutler *et al.* (1998) points out that there is great reluctance to put a monetary value on lives saved or extended and yet such a valuation is necessary to measure the price change $p_W^t/p_W^{t-1}$. He suggests the direct use of cost-effectiveness studies in health care by using an adjustment factor that is based on Quality-Adjusted Life Years (QALY). Christian (2007) applies an adjustment of this kind to a volume index of hospital services and finds a significant positive effect on the measured rate of output.

[18]"Analyses aggregated from treatments clearly shown to be medically effective suggest that at least half the life-expectancy gains since 1950 are due to medical advances. About 90 percent of the gains in life expectancy are attributable to improvements in the rates of death in infancy and the rates of death from cardiovascular disease. Prevailing estimates suggest that at least half the reduction in these rates are due to medical care. We therefore assumed in our base case that 50 percent of the total gains in life expectancy were due to medical care" (Cutler *et al.*, 2006, p. 921).

of the product. It would seem that a measure of service production should not be influenced by the *individual* capacity of the consumer to make use of these services. For example, the same teaching activity performed on two different students should be measured as the same quantity of teaching services toward each student, even if it turns out that one of them benefits less from teaching than the other. Or the same medical treatment, applied to two different persons with the same disease (and similar patient characteristics), should be measured as equal quantities of medical services. Unless, that is, the two persons come from different groups of patients, where "groups" is understood as sets of patients with characteristics that require *a priori* different services, for example young and old persons for particular treatments. It is easy to see that making this difference in practice will be difficult and sometimes impossible. This limits the applicability of output methods that rely on the direct contribution to outcomes.

## 6. Conclusions

This paper looked at the notions of output and outcome in the context of service activities with market and non-market producers. We started by defining the various notions, in particular inputs, processes/activities, outputs, and outcomes. The following main conclusions were derived.

Even in the simplest market model without any quality change in products, output measurement may require some, often implicit, reference to outcome when products are grouped and classified. Thus, output and outcome are different, but they are not independent of each other.

With non-market producers, the equality of marginal utility and unit cost ratios breaks down. The money value of output is determined as the sum of costs of inputs, but it is possible to construct meaningful indices of cost-based volumes and quasi prices along with indices of volumes and prices of inputs. Akin to the simple market case, cost-based indices are not void of implicit information about utility and outcome—typically, this enters via the product classification on the basis of which aggregation takes place.

In the presence of quality change, all existing methods require implicit or explicit information or reasoning about outcome. The simplest and most widely-used way of controlling for quality change—matching or stratification—requires implicit statements about the similarity of products in their capacity to generate outcome when products are grouped together. A second method uses explicit information on product characteristics to adjust unit costs for differences in characteristics. The choice of characteristics cannot proceed without considering outcome.

Problems of quality adjustment arise whether services are provided by market or non-market producers. The fact that there is an observable market transaction in one period and another market transaction in the next does not imply that they are comparable—otherwise, price statisticians would be in the convenient situation of only having to deal with quality adjustment for non-market production. This is of course not the case. Although the distinction between market and non-market producers is useful and has consequences for measuring the current price value of output, it loses most of its significance when we come to the thorny part of measuring output—the treatment of quality change and the treatment of exiting and new products.

A pragmatic approach will be called for to proceed with services measurement. In particular, there is no reason to approach every type of service with the same method for quality adjustment—methodologies should be robust but will also have to be reflective of data availability and transparency for users.

Measuring output for complex services is difficult, but the conclusion should not be that it is simply too difficult to do anything. Health and education account for a too large and growing part of the economy to ignore output measurement for them. It may take a while before consensual and internationally comparable methods are agreed upon, but active research and data development is vital to achieve this objective.

## References

Abraham, K. G. and Ch. Mackie, "A Framework for Nonmarket Accounting," in D. Jorgenson, J. S. Landefeld, and W. Nordhaus (eds), *A New Architecture for the U.S. National Account, NBER, Studies in Income and Wealth*, University of Chicago Press, Chicago, 161–92, 2006.

Aigner, D. J., C. A. K. Lovell, and P. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, 21–37, 1977.

Archibald, R. B., "On the Theory of Industrial Price Measurement: Output Price Indexes," *Annals of Economic and Social Measurement*, 6, 57–72, 1977.

Armesto, S. G., M. L. Lapetra, L. Wei, and E. Kelley, "Health Care Quality Indicators Project 2006: Data Collection Update Report," OECD Health Working Papers 29, Paris, 2007.

Atkinson, A. B., *Atkinson Review. Final Report: Measurement of Government Output and Productivity for the National Accounts*, Palgrave Macmillan, New York, 2005.

Balk, B. M., *Industrial Price, Quantity and Productivity Indices, The Micro-Economic Theory and an Applicati*on, Kluwer Academic Publishers, Boston, 1998.

Berndt, E. R., D. M. Cutler, R. G. Frank, Z. Griliches, J. P. Newhouse, and J. E. Triplett, "Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues," NBER Working Paper Series, No. 6817, 1998.

Christian, M. S., "Measuring the Output of Health Care in the United States," *Survey of Current Business*, June, 78–83, 2007.

Court, L. M., "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, General Motors Corporation, New York, 99–117, 1939.

Cutler, D. M., M. McClellan, J. P. Newhouse, and D. Remler, "Are Medical Prices Declining? Evidence from Heart Attack Treatment," *Quarterly Journal of Economics*, CXII, 991–1023, 1998.

Cutler, D. M. and E. R. Berndt (eds), *Medical Care Output and Productivity*, Vol. 62, NBER Studies in Income and Wealth, University of Chicago Press, Chicago, 2001.

Cutler, D. M., A. B. Rosen, and S. Vijan, "The Value of Medical Spending in the United States, 1960–2000," *New England Journal of Medicine*, 355, 920–7, 2006.

Diewert, W. E., "Application of Duality Theory," in M. D. Intriligator and D. A. Kendrick (eds), *Frontiers of Quantitative Economics*, Vol. II, North Holland, Amsterdam, 106–71, 1974.

———, "The Measurement of Nonmarket Sector Outputs and Inputs Using Cost Weights," Discussion Paper 08-03, Department of Economics, University of British Columbia, January 17, 2008 (available at http://www.econ.ubc.ca/diewert/dp0803.pdf).

Diewert, W. E. and A. O. Nakamura, "The Measurement of Productivity for Nations," in J. J. Heckman and E. E. Leamer (eds), *Handbook of Econometrics*, Volume 6A, Elsevier, 4502–86, 2007.

Eurostat, *Handbook on Price and Volume Measures in National Accounts*, European Communities, Luxembourg, 2001.

Färe, R., S. Grosskopf, and C. A. K. Lovell, *The Measurement of Efficiency of Production*, Kluwer Academic Publishers, Boston, 1985.

Farrell, M. J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, CXX, part 3, 253–90, 1957.

Fisher, F. M. and K. Shell, *The Economic Theory of Price Indices*, Academic Press, New York and London, 1972.
Gold, M. R., J. E. Siegel, L. B. Russel, and M. C. Weinstein (eds), *Cost Effectiveness in Health and Medicine*, Oxford University Press, New York, 1996.
Griliches, Z. (ed.), *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Harvard University Press, Cambridge, MA, 1971.
Hill, T. P., *Price and Volume Measures for Non-Market Services*, Report to the Statistical Office of the European Communities, Brussels, 1975.
International Monetary Fund, International Labour Organisation, Organisation for Economic Co-operation and Development, United Nations Commission for Europe, and The World Bank, *Producer Price Index Manual, Theory and Practice*, Washington DC, 2004.
Jorgenson, Dale W. and Barbara Fraumeni, "The Accumulation of Human and Nonhuman Capital, 1948–84," in R. E. Lipsey and H. S. Tice (eds), *The Measurement of Saving, Investment and Wealth*, National Bureau of Economic Research Studies in Income and Wealth, Vol. 52, University of Chicago Press, Chicago, 227–82, 1989.
Koechlin, F., L. Lorenzoni, and P. Schreyer, "Comparing Price Levels of Hospital Services Across Countries: Results of a Pilot Study," OECD Statistics Directorate Working Paper No. 32, OECD, Paris, 2010.
Lawrence, D. and E. W. Diewert, "Regulating Electricity Networks: The ABS of Setting X in New Zealand," in T. Coelli and D. Lawrence (eds), *Performance Measurement and Regulation of Network Utilities*, Edward Elgar Publishing, Cheltenham, 207–41, 2006.
Pakes, A., "A Reconsideration of Hedonic Price Indexes with an Application to PCs," *American Economic Review*, 93, 1578–96, 2003.
Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55, 1974.
Rosen, A. B. and D. Cutler, "Measuring Medical Care Productivity: A Proposal for U.S. National Health Accounts," *Survey of Current Business*, June, 54–8, 2007.
Schreyer, P., "Towards Measuring the Volume Output of Education and Health Services: A Handbook," OECD Statistics Directorate Working Paper No. 31, Paris, 2010.
Shephard, R. W., *Theory of Cost and Production Functions*, Princeton University Press, Princeton, 1970.
Stone, R., *Quantity and Price Indexes in National Accounts*, Organisation for European Economic Cooperation, Paris, 1956.
Triplett, J. E., "What's Different about Health? Human Repair and Car Repair in National Accounts and in National Health Accounts," in D. Cutler and E. Berndt (eds), *Medical Care Output and Productivity*, NBER Studies in Income and Wealth, Volume 62, University of Chicago Press, Chicago, 1998.
———, *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD, Paris, 2006.
Triplett, J. E. and B. P. Bosworth, *Productivity in the U.S. Services Sector: New Sources of Economic Growth*; The Brookings Institution, Washington DC, 2004.
Triplett, J. E. and R. J. McDonald, "Assessing the Quality Error in Output Measures: The Case of Refrigerators," *Review of Income and Wealth*, 23, 137–56, 1977.
Vanoli, A., *Une Histoire de la Comptabilité Nationale*, La Découverte, Paris, 2002.