The Measurement of Nonmarket Sector Outputs and Inputs Using Cost Weights

W. Erwin Diewert,[1]
Revised March 15, 2008.
Discussion Paper 08-03,
Department of Economics,
University of British Columbia,
Vancouver, B.C.,
Canada, V6T 1Z1.
Email: diewert@econ.ubc.ca

Abstract

In many sectors of the economy, governments either provide various services at no cost or at highly subsidized prices. Examples are the health, education and general government sectors. The *System of National Accounts 1993* recommends valuing these nonmarket outputs at their costs of production but it does not give much guidance on exactly how to do this. In this paper, an explicit methodology is developed that enables one to construct these marginal cost prices. However, in the main text, an activity analysis approach is taken in order to simplify the analysis, so in particular, constant returns to scale, no substitution production functions for the specific activities in the nonmarket sector are assumed. It is shown that it is possible to obtain meaningful measures of Total Factor Productivity growth in this framework. An Appendix relaxes some of the restrictive assumptions that are used in the main text.

Keywords

Measurement of output, input and productivity, nonmarket sector, health, education, general government, cost functions, duality theory, marginal cost prices, activity analysis, technical progress, total factor productivity, outlet substitution bias.

JEL Classification Numbers

C43, D24, D57, E23, H40, H51, H52, I10, I20, O47.

1. Introduction

In this paper, we will examine exactly how outputs can be measured in industries such as the health sector where reliable final demand prices are not available. The existing national income accounting methodology suggests using cost weights as prices in this situation[2] and in the main text, we will work out the algebra for implementing this

---

[1] The author thanks Ian Bobbin, Kevin Fox, Peter Hill, Denis Lawrence, Carl Obst, Paul Schreyer, Mick Silver and Kam Yu for helpful discussions on this topic and the Australian Research Council and the SSHRC of Canada for financial support. None of the above are responsible for any opinions expressed in the paper.

[2] The literature on this topic dates back to Hicks (1940).

methodology under somewhat simplified assumptions; i.e., we will assume that each output is produced using a constant returns to scale, no substitution type production function. However, we will allow for technical progress; i.e., we will allow for the introduction of new processes that deliver the same services using fewer resources.

In section 2 below, we consider the case of two processes, activities or medical procedures and construct "industry" measures of output, input and productivity using unit costs as weights. We find that it is possible to have productivity growth in this context.[3]

In section 3, we indicate that the methodology developed in section 2 can break down under certain conditions. Thus when a new procedure is introduced that has an output that is equivalent to the output of an existing procedure, the section 2 methodology will give a rate of real output growth that is too small. We indicate how this problem can be addressed.

Section 4 concludes.

An Appendix reworks the methodology when the assumption of no substitution production functions is relaxed.

2. The Measurement of Cost Weighted Outputs Using Fixed Coefficient Technologies

In order to minimize notational complexities, we will consider only the case of two medical *procedures* such as performing an operation on a patient with reasonably homogeneous characteristics or diagnosing a medical condition and prescribing a treatment.[4] Let the number of type i procedures performed in period t by a particular establishment be y_i^t for i = 1,2 and t = 0,1. When a type i procedure is performed in period t, there is a *vector of input requirements*, $a_i^t \equiv [a_{i1}^t,a_{i2}^t,...,a_{iN(i)}^t]$ for i = 1,2, associated with each procedure where N(i) is the number of inputs used by procedure i.[5] We assume a constant returns to scale, fixed coefficients production function for each procedure[6] and so the *vector of inputs* used by procedure i and period t, $x_i^t \equiv [x_{i1}^t,x_{i2}^t,...,x_{iN(i)}^t]$, is equal to the period t input-output vector for procedure i, a_i^t, times the *number of procedures* of type i performed in period t, y_i^t; i.e., we have the following

[3] When measuring the output of nonmarket sectors, statistical agencies often assume that output growth is equal to input growth and hence Total Factor Productivity (TFP) growth is not possible using this methodology. However, following the work of Mai (2004), Pritchard (2004) and Atkinson (2005) in the UK, we will show that if statistical agencies can compute input requirements per unit of output in these nonmarket sectors, then using unit costs as price weights for outputs can lead to output growth rates that are faster than the corresponding input growth rates and hence TFP improvements are possible in this environment.

[4] The same analysis can be applied to any nonmarket service where there are outputs that can be measured in reasonably homogeneous quantity units. For a more realistic discussion of some of the problems that arise when measuring health sector outputs, see Yu and Ariste (2008).

[5] The vector a_i^t is the vector of *input-output coefficients* for the ith technology in period t. Note that the inputs include both intermediate and primary inputs.

[6] These assumptions are somewhat problematic due to the existence of fixed costs in hospitals. There are also problems associated with the allocation of overhead costs.

relationship between the amounts of inputs used by procedure i in period t, x_i^t, and the output produced by procedure i in period t, y_i^t:

$$(1)\ x_i^t = a_i^t\, y_i^t\ ; \qquad i = 1,2\ ;\ t = 0,1.$$

In each period t, there is a vector of positive *input prices*, $w_i^t \equiv [w_{i1}^t, w_{i2}^t, ..., w_{iN(i)}^t]$ for i = 1,2, associated with each procedure and the *period t cost associated with performing type i procedures*, C_i^t, is simply the sum of the product of the individual input prices w_{in}^t times the amounts of input used in period t, x_{in}^t:

$$(2)\ C_i^t \equiv w_i^t \cdot x_i^t \equiv \sum_{n=1}^{N(i)} w_{in}^t\, x_{in}^t\ ; \qquad i = 1,2\ ;\ t = 0,1.$$

Note that the above framework allows for the possibility of *technical progress* in the delivery of each procedure going from period 0 to period 1; i.e., we have *not* assumed that a_i^0 equals a_i^1 for each procedure i or in words, we have not assumed that the vector of input-output coefficients remains constant for each procedure. If service providers in period 1 have access to the same technology as was used in period 0, then the following inequalities will be satisfied:

$$(3)\ w_i^1 \cdot a_i^1 \leq w_i^1 \cdot a_i^0\ ; \qquad i = 1,2.$$

However, if there is *technical progress* for procedure i going from period 0 to 1, then the weak inequality in (3) for this i will hold strictly, so that the unit cost of delivering the ith type of service, $w_i^1 \cdot a_i^1$, *goes down* if we use the new technology characterized by the input output vector a_i^1 compared to the unit cost of using the old technology, which is $w_i^1 \cdot a_i^0$.[7]

Thus if there is *no technical progress* going from period 0 to 1 in both procedures, the input-output vectors remain constant so that we have:

$$(4)\ a_i^0 = a_i^1\ ; \qquad i = 1,2.$$

If there is *technical progress* in both procedures going from period 0 to 1, then the following inequalities hold:

$$(5)\ w_i^1 \cdot a_i^1 < w_i^1 \cdot a_i^0\ ; \qquad i = 1,2.$$

We will draw on assumptions (4) and (5) subsequently.

Suppose now that we define *health industry aggregates* that aggregate together the inputs and outputs of the two activities described above. It can be seen that it is straightforward

[7] This definition of technical progress for sector i is weaker than the alternative definition that the vector of period 1 input-output coefficients a_i^1 be equal to or less than the corresponding period 0 vector of input-output coefficients a_i^0 with a strict inequality for at least one component; i.e., the alternative (strong) definition of technical progress for sector i is $a_i^1 < a_i^0$ where $a_i^1 < a_i^0$ means $a_i^1 \leq a_i^0$ but $a_i^1 \neq a_i^0$. Our weaker definition seems to be a more suitable one for the present purpose.

to construct aggregate price and quantity indexes for inputs. Thus the *Laspeyres industry input quantity index* Q_L^* can be defined as follows:[8]

$$(6)\ Q_L^* \equiv [w_1^0 \cdot x_1^1 + w_2^0 \cdot x_2^1]/[w_1^0 \cdot x_1^0 + w_2^0 \cdot x_2^0]$$
$$= [w_1^0 \cdot a_1^1 y_1^1 + w_2^0 \cdot a_2^1 y_2^1]/[w_1^0 \cdot a_1^0 y_1^0 + w_2^0 \cdot a_2^0 y_2^0] \qquad \text{using (1).}$$

Similarly, the *Paasche and Fisher industry input quantity indexes*, Q_P^* and Q_F^*, can be defined as follows:

$$(7)\ Q_P^* \equiv [w_1^1 \cdot x_1^1 + w_2^1 \cdot x_2^1]/[w_1^1 \cdot x_1^0 + w_2^1 \cdot x_2^0]$$
$$= [w_1^1 \cdot a_1^1 y_1^1 + w_2^1 \cdot a_2^1 y_2^1]/[w_1^1 \cdot a_1^0 y_1^0 + w_2^1 \cdot a_2^0 y_2^0] \qquad \text{using (1);}$$

$$(8)\ Q_F^* \equiv [Q_L^* Q_P^*]^{1/2} .$$

Turning now to the corresponding problem of defining industry output aggregates, we encounter a severe problem: namely, although we can observe the output quantities for each procedure (the y_i^t), we generally will not be able to observe the corresponding output prices.[9] Without output prices, normal index number theory cannot be applied. For many purposes (including the measurement of welfare), the desired conceptual price for each type of medical service is a *household marginal valuation price* or a *final demand price*; i.e., the price that a household would be willing to pay for an extra unit of the service. But it is difficult for experts to agree on what the appropriate final demand prices should be in the context of pricing medical services. If experts cannot agree, this puts statistical agencies in a difficult position since their estimates of output and input should be *objective* and *reproducible*.

Given that final demand prices are generally not available, the *System of National Accounts 1993* recommends valuing publicly provided services at their costs of production.[10] Thus in the remainder of this note, we will follow this advice and examine the output indexes which are implied by following this recommendation in the context of our very simple model.

[8] For the definitions of the Laspeyres, Paasche and Fisher ideal price indexes, see Fisher (1922). The corresponding quantity indexes can be obtained from the same formulae but with the roles of prices and quantities reversed.
[9] See Atkinson (2005; 88-90) for a nice discussion on the valuation of nonmarket outputs and the differences between marginal cost and final demander valuations.
[10] In particular, Chapter 16 in *SNA 1993* notes that if we have quantity information on the numbers of various different types of tightly specified medical procedures, then Laspeyres or Paasche indexes can be calculated using sales as weights for market services and costs for nonmarket services. The situation is summarized in paragraphs 16.133 and 16.134 of the 1993 SNA on non-market goods and services which was written by Peter Hill. Paragraph 16.134 says: " In principle, volume measures may be compiled directly by calculating a weighted average of the quantity relatives for the various goods or services produced as outputs using the values of these goods and services as weights. Exactly the same method may be applied even when the output values have to be estimated on the basis of their costs of production."

The *unit cost of production for procedure i in period t*, p_i^t, can readily be defined in our simple model as the cost of producing one unit of the procedure using the period i technology:

(9) $p_i^t \equiv w_i^t \cdot a_i^t$; $\quad i = 1,2$; $t = 0,1$.

Using the above cost based procedure i output prices p_i^t along with the corresponding period t output quantities y_i^t, we can readily define the *Laspeyres, Paasche and Fisher output quantity indexes*, Q_L, Q_P and Q_F respectively:[11]

(10) $Q_L \equiv [p_1^0 y_1^1 + p_2^0 y_2^1]/[p_1^0 y_1^0 + p_2^0 y_2^0]$
$= [w_1^0 \cdot a_1^0 y_1^1 + w_2^0 \cdot a_2^0 y_2^1]/[w_1^0 \cdot a_1^0 y_1^0 + w_2^0 \cdot a_2^0 y_2^0]$ using (9);

(11) $Q_P \equiv [p_1^1 y_1^1 + p_2^1 y_2^1]/[p_1^1 y_1^0 + p_2^1 y_2^0]$
$= [w_1^1 \cdot a_1^1 y_1^1 + w_2^1 \cdot a_2^1 y_2^1]/[w_1^1 \cdot a_1^1 y_1^0 + w_2^1 \cdot a_2^1 y_2^0]$ using (9);

(12) $Q_F \equiv [Q_L Q_P]^{1/2}$.

Comparing the Laspeyres and Paasche output indexes (10) and (11) with their input counterparts (6) and (7) leads to some interesting results when we make the assumption of no technical change in the procedures, Assumption (4), or when we make the assumption of technical change in both procedures, Assumption (5); i.e., we can prove the following two Propositions:

Proposition 1: If there is no technical progress going from period 0 to 1 in each procedure, then the Laspeyres, Paasche and Fisher output indexes are exactly equal to the corresponding Laspeyres, Paasche and Fisher input indexes; i.e., we have:

(13) $Q_L = Q_L^*$;
(14) $Q_P = Q_P^*$;
(15) $Q_F = Q_F^*$.

Proof: If (4) holds, then the second equation in (6) is equal to the second equation in (10) which establishes (13). Similarly if (4) holds, then the second equation in (7) is equal to the second equation in (11) which establishes (14). Finally, (15) follows from (13) and (14). Q.E.D.

If we define *productivity growth* as an output index divided by an input index, then the above Proposition tells us that there will be no Laspeyres, Paasche or Fisher productivity growth in the industry if there is no technological progress in the procedure technologies

[11] Cost weighted Laspeyres type output quantity indexes of the type defined by (10) are used widely in the UK in recent years when constructing measures of nonmarket output quantity growth; see Mai (2004; 65), Pritchard (2004; 78) and Atkinson (2005; 88).

and if we use the same index number formula (Laspeyres, Paasche or Fisher) for measuring both input and output growth.[12]

Proposition 2: If there is technical progress going from period 0 to 1 in each procedure,[13] then the Laspeyres, Paasche and Fisher output indexes are strictly greater than the corresponding Laspeyres, Paasche and Fisher input indexes; i.e., we have:

(16) $Q_L > Q_L^*$;
(17) $Q_P > Q_P^*$;
(18) $Q_F > Q_F^*$.

Proof: If (5) holds, then the second equation in (6) is strictly less than the second equation in (10) which establishes (16). Similarly if (5) holds, then the second equation in (7) is strictly less than the second equation in (11) which establishes (17). Finally, (18) follows from (16) and (17). Q.E.D.

Thus if there is technical progress in either procedure, then the above Proposition tells us that there will be Laspeyres, Paasche or Fisher productivity growth in the industry if we use the same index number formula (Laspeyres, Paasche or Fisher) for measuring both input and output growth. This is a somewhat important result because it is sometimes thought that using cost weights to price outputs in nonmarket sectors like health, education and general government leads to output growth measures that are equal to input growth measures and so that productivity improvements in these hard to measure sectors must be nonexistent using *SNA 1993* methodology.[14] Proposition 2 demonstrates that *there can be productivity improvements that will show up using cost based prices*, provided that we can capture any technological improvements in the delivery of these hard to measure services by accurately estimating the input-output coefficients for the delivery of one unit of the service in each period.[15]

---

[12] In the Appendix, we show that when we relax the fixed coefficients assumption we are using here, the use of the Fisher index is clearly preferred over its Paasche and Laspeyres counterparts.
[13] It is easy to modify the Proposition and obtain the same results if we have no technical progress in one procedure but strict progress in the other.
[14] Of course, this statement is not true as was noted above when the work of Mai (2004), Pritchard (2004) and Atkinson (2005) in the UK was cited.
[15] If it proves to be difficult or impossible to measure nonmarket output quantities, then economic statisticians have generally measured the value of nonmarket outputs by the value of inputs used and implicitly or explicitly set the price of nonmarket output equal to the corresponding nonmarket input price index. Atkinson (2005; 12) describes the situation in the UK prior to 1998 as follows: “In many countries, and in the United Kingdom from the early 1960’s to 1998, the output of the government sector has been measured by convention as the value equal to the total value of inputs; by extension the volume of output has been measured by the volume of inputs. This convention regarding the volume of government output is referred to below as the (output = input) convention, and is contrasted with direct measures of government output. The inputs taken into account in recent years in the United Kingdom are the compensation of employees, the procurement costs of goods and services and a charge for the consumption of fixed capital. In earlier years and in other countries, including the United States, the inputs were limited to employment.” Note that the above conventions imply that capital services input for government owned capital will generally be less than the corresponding capital services input if the capital services were rented or leased. In the owned case, the government user cost of capital consists only of depreciation but in the leased case, the rental rate would cover the cost of depreciation plus the opportunity cost of the financial capital tied up

It can be seen that the restriction to only two procedures is not material in the above analysis; it can readily be generalized to an arbitrary number of procedures.

Up to this point, it appears that setting prices equal to average costs[16] is a reasonable strategy when pricing medical service outputs. However, in the following section, we show that this methodology does not give reasonable results when a new medical technology is developed that is more efficient than an existing technology but the new technology does not immediately displace the old technology.

3. The Introduction of a New More Efficient Technology

We now suppose that procedure 1 is a well established "incumbent" procedure which has input-output coefficient vectors a_1^0 and a_1^1 in periods 0 and 1 respectively but that procedure 2 is a "new" improved procedure that accomplishes exactly what procedure 1 accomplishes but it is only available in period 1.[17] The vector of input-output coefficients for procedure 2 in period 1 is a_2^1. The assumption that the new procedure is more efficient than the old procedure in period 1 means that one unit of the new procedure has a lower unit cost than one unit of the old procedure in period 1 so that

(19) $w_2^1 \cdot a_2^1 < w_1^1 \cdot a_1^1$.

We continue to assume that the inequality (3) holds for the incumbent technology; i.e. (3) holds for i = 1. Now we can use the period 1 input-output coefficients for the new technology as *imputed* input-output coefficients for period 0 but of course, the corresponding period 0 output and input for the new procedure, y_2^0 and x_2^0, should be set equal to zero; i.e., we make the following assumptions:

(20) $a_2^0 \equiv a_2^1$; $y_2^0 \equiv 0$; $x_2^0 \equiv 0$.

We assume that the new procedure is also more efficient than the old procedure using the input prices of period 0; i.e., we assume that:[18]

(21) $w_2^0 \cdot a_2^0 < w_1^0 \cdot a_1^0$.

---

in the capital input. Atkinson (2005; 49) makes the following recommendation on this issue: "We recommend that the appropriate measure of capital input for production and productivity analysis is the flow of *capital services* of an asset type. This involves adding to the capital consumption an interest charge, with an agreed interest rate, on the entire owned capital." We concur with Atkinson's recommendation.

[16] Under our constant returns to scale assumptions being used here, these average costs are also equal to marginal costs.

[17] Paul Schreyer verbally suggested the unit value type methodology that is developed in this section.

[18] There is another implicit assumption here; i.e., we are assuming that it is possible to obtain estimates of the prices of the inputs that could have been used in period 0 if the new technology were available in period 0. Thus we are assuming that it is possible to obtain estimates of the period 0 input price vector w_2^0 for the new technology on a retrospective basis for period 0.

The methodology developed in the previous section still seems to be satisfactory if we are constructing input quantity indexes. However, the previous section methodology no longer seems to be satisfactory when constructing output quantity indexes. The problem with the old methodology is this: under our assumption that the output of the new procedure is completely equivalent to the output of the old procedure, it is easy to see that the "correct" output growth index should be the following *unit value quantity index*, Q_{UV}:

(22) $Q_{UV} \equiv (y_1^1 + y_2^1)/y_1^0$.

However, if we compute the Laspeyres output quantity index Q_L defined by (10) under our current assumptions, we find that:

$$
\begin{aligned}
(23)\ Q_L &\equiv [p_1^0 y_1^1 + p_2^0 y_2^1]/[p_1^0 y_1^0 + p_2^0 y_2^0] && \\
&= [w_1^0 \cdot a_1^0 y_1^1 + w_2^0 \cdot a_2^0 y_2^1]/[w_1^0 \cdot a_1^0 y_1^0 + w_2^0 \cdot a_2^0 y_2^0] && \text{using (9);} \\
&= [w_1^0 \cdot a_1^0 y_1^1 + w_2^0 \cdot a_2^0 y_2^1]/w_1^0 \cdot a_1^0 y_1^0 && \text{using (20); i.e., } y_2^0 = 0 \\
&= [y_1^1/y_1^0] + [w_2^0 \cdot a_2^0/w_1^0 \cdot a_1^0][y_2^1/y_1^0] && \text{rearranging terms} \\
&< [y_1^1/y_1^0] + [y_2^1/y_1^0] && \text{using (21)} \\
&= Q_{UV} && \text{using definition (22).}
\end{aligned}
$$

Similarly, if we compute the Paasche output quantity index Q_P defined by (11) under our current assumptions, we find that:

$$
\begin{aligned}
(24)\ Q_P &\equiv [p_1^1 y_1^1 + p_2^1 y_2^1]/[p_1^1 y_1^0 + p_2^1 y_2^0] && \\
&= [w_1^1 \cdot a_1^1 y_1^1 + w_2^1 \cdot a_2^1 y_2^1]/[w_1^1 \cdot a_1^1 y_1^0 + w_2^1 \cdot a_2^1 y_2^0] && \text{using (9);} \\
&= [w_1^1 \cdot a_1^1 y_1^1 + w_2^1 \cdot a_2^1 y_2^1]/w_1^1 \cdot a_1^1 y_1^0 && \text{using (20); i.e., } y_2^0 = 0 \\
&= [y_1^1/y_1^0] + [w_2^1 \cdot a_2^1/w_1^1 \cdot a_1^1][y_2^1/y_1^0] && \text{rearranging terms} \\
&< [y_1^1/y_1^0] + [y_2^1/y_1^0] && \text{using (19)} \\
&= Q_{UV} && \text{using definition (22).}
\end{aligned}
$$

But (23) and (24) imply the following result as well:

(25) $Q_F < Q_{UV}$.

Thus if the new procedure has an output that is equivalent to the existing procedure, then use of the cost based output price methodology developed in the previous section will give Laspeyres, Paasche and Fisher output growth indexes that are biased downwards compared to the conceptually correct unit value output growth index defined by (22). Hence under the assumptions of this section, *the methodology developed in the previous section will lead to estimates of output growth and productivity growth that are biased downwards*.

The bias that results from the incorrect measurement of the effects of the introduction of a new and more efficient procedure is similar in many respects to *new outlets bias*; i.e., a new output is linked into an index in such a way that the index shows no change when in fact, a change should be recorded.[19] The methodology for dealing with a new procedure developed above is relatively straightforward and is reproducible and objective under the assumptions of the model. However, in the real world, a new procedure is unlikely to have an output that is *exactly* equivalent to the output of an existing procedure and hence in real life, it will not be so straightforward to deal with new medical procedures.

4. Conclusion

The valuation of outputs produced by the nonmarket sector is a complicated task. A first best solution would be to have unambiguous, objective and reproducible final demand prices but this solution is generally not available to statistical agencies. A second best solution is to value outputs at their average costs as is recommended in *SNA 1993*.[20] We have developed the algebra associated with this second approach and found (not surprisingly) that it is possible to have Total Factor Productivity growth using this methodology, provided that sufficient information is available.

It should be noted that the methodology developed here is applicable to a wide variety of industries that are either entirely nonmarket, or partially nonmarket, as in the case of regulated industries.[21]

Appendix: The Measurement of Output, Input and Productivity in the Nonmarket Sector for More General Production Functions

In this Appendix, we will relax the assumption that the procedure production functions are of the fixed coefficients, no input substitution variety. Theoretical output, input and productivity indexes will be defined in this more general context and we will exhibit various observable indexes that can approximate these theoretical indexes to the accuracy of at least a first order approximation.

As in section 2, we assume that there are two procedures in periods 0 and 1 but now we assume that there is a *production function*, f_i^t, for procedure i in period t where $y_i^t = f_i^t(x_{i1}^t,x_{i2}^t,...,x_{iN(i)}^t) = f_i^t(x_i^t)$ is the amount of output for procedure i that can be produced by the input vector x_i^t in period t for t = 0,1 and i = 1,2. We assume that each production function $f_i^t(x_i)$ is a nonnegative, increasing, continuous, concave and linearly homogeneous function in the components of its input vector x_i. As in section 2, we assume that in each period t, producers in sector i face the positive vector of input prices, $w_i^t \equiv [w_{i1}^t,w_{i2}^t,...,w_{iN(i)}^t]$ for i = 1,2 and t = 0,1. For each period t and each sector i, the sector i *total cost function* $C_i^t(y_i,w_i)$ associated with each procedure can be defined as follows:

[19] For discussions on how to account for outlet substitution bias in the context of a consumer price index, see Diewert (1995) (1998) and Chapter 11 in the ILO (2004).
[20] See sections 16.133-16.135 in the *SNA 1993*.
[21] With the growth of incentive type regulatory regimes, there is increasing interest in forming output aggregates using marginal cost weights for prices in order to calculate the Total Factor Productivity growth of regulated firms; see Lawrence and Diewert (2006) for an extensive discussion of the issues. It should be noted that Lawrence and Diewert assumed separable technologies of the type considered in this paper.

$$(A1)\ C_i^t(y_i,w_i) \equiv \min_{x's} \{w_i \cdot x_i : f_i^t(x_i) \geq y_i\} ; \qquad i = 1,2 ; t = 0,1$$
$$= c_i^t(w_i)y_i$$

where $c_i^t(w_i) \equiv C_i^t(1,w_i)$ is the period t *unit cost function* for sector i; i.e., it is equal to the minimum cost of producing one unit of sector i output using the period t technology if the sector faces the vector of input prices w_i. The unit cost function c_i^t will satisfy the same regularity conditions as the production function f_i^t; i.e., $c_i^t(w_i)$ will be a nonnegative, increasing, continuous, concave and linearly homogeneous function in the components of the input price vector w_i.[22]

We assume that in each period, producers minimize the cost of producing their procedure outputs.[23] Thus letting y_i^t and x_i^t denote the observed scalar output and input vector of sector i in period t, we will have the following equalities:

$$(A2)\ w_i^t \cdot x_i^t = C_i^t(y_i^t,w_i^t) = c_i^t(w_i^t)y_i^t ; \qquad i = 1,2 ; t = 0,1.$$

We also assume that each unit cost function is differentiable at the observed input prices for each sector and each period so that Shephard's (1953; 11) Lemma implies the following relationships between the input quantity vectors x_i^t and the corresponding output levels y_i^t:

$$(A3)\ x_i^t = \nabla_w C_i^t(y_i^t,w_i^t) = \nabla_w c_i^t(w_i^t)y_i^t ; \qquad i = 1,2 ; t = 0,1.$$

The observed input-output vectors a_i^t for each sector i and each time period can be defined as the observed input vectors x_i^t divided by the corresponding output levels y_i^t:

$$(A4)\ a_i^t \equiv x_i^t/y_i^t ; \qquad i = 1,2 ; t = 0,1.$$

Comparing (A3) with (A4) shows that the vectors of first order partial derivatives of the unit cost functions, $\nabla_w c_i^t(w_i^t)$, are also observable and are equal to the corresponding input-output vectors a_i^t:

$$(A5)\ \nabla_w c_i^t(w_i^t) = a_i^t ; \qquad i = 1,2 ; t = 0,1.$$

The period t unit cost for sector i, $c_i^t(w_i^t)$, are also observable and can serve as our *cost based output prices* p_i^t for units of output in sector i during period t; i.e., define the output prices p_i^t as follows:

$$(A6)\ p_i^t \equiv c_i^t(w_i^t) = w_i^t \cdot a_i^t = w_i^t \cdot x_i^t/y_i^t ; \qquad i = 1,2 ; t = 0,1.$$

[22] For background information on cost and production functions and their regularity conditions, see Diewert (1974).

[23] Obviously, in many situations where governments are in charge of producing the procedure outputs, this assumption will not be satisfied. However, in order to make some progress on our index number problems, we will make this assumption.

Definitions (A6) imply the following relationships between the value of output $p_i^t y_i^t$ and the value of input $w_i^t \cdot x_i^t$ in sector i for period t:

$$\text{(A7)}\quad p_i^t y_i^t = c_i^t(w_i^t) y_i^t = w_i^t \cdot x_i^t\ ; \qquad i = 1,2\ ;\ t = 0,1.$$

We now in a position to define the health sector's period t total cost function, C^t, but first we require some further notation. Let $y \equiv [y_1,y_2]$ be a two dimensional reference vector of possible health sector outputs and let $w \equiv [w_1,w_2]$ be an N(1)+N(2) dimensional vector of reference input prices. Define the *health sector's period t total cost function* C^t as the sum of the period t procedure cost functions:

$$\begin{aligned} \text{(A8)}\quad C^t(y,w) &\equiv C_1^t(y_1,w_1) + C_2^t(y_2,w_2)\ ; && t = 0,1 \\ &= c_1^t(w_1)y_1 + c_2^t(w_2)y_2 && \text{using (A2).} \end{aligned}$$

We will use the sector's total cost function $C^t(y,w)$ in order to define indexes of health sector technical progress, output growth and input price growth going from period 0 to period 1 in what follows.[24] As a preliminary step, insert the data pertaining to period t into definition (A8) and we obtain the following equations:

$$\begin{aligned} \text{(A9)}\quad C^t(y^t,w^t) &= c_1^t(w_1^t)y_1^t + c_2^t(w_2^t)y_2^t\ ; && t = 0,1 \\ &= w_1^t \cdot x_1^t + w_2^t \cdot x_2^t && \text{using (A7)} \\ &= p_1^t y_1^t + p_2^t y_2^t && \text{using (A6).} \end{aligned}$$

We now use the total cost function in order to define a *family of cost based output quantity indexes*, $\alpha(y^0,y^1,w,t)$, as follows:

$$\begin{aligned} \text{(A10)}\quad \alpha(y^0,y^1,w,t) &\equiv C^t(y^1,w)/C^t(y^0,w) \\ &= [c_1^t(w_1)y_1^1 + c_2^t(w_2)y_2^1]/[c_1^t(w_1)y_1^0 + c_2^t(w_2)y_2^0] && \text{using (A8).} \end{aligned}$$

Thus the theoretical output quantity index $\alpha(y^0,y^1,w,t)$ defined by (A10) is equal to the (hypothetical) total cost $C^t(y^1,w)$ of producing the vector of observed period 1 procedure outputs, $y^1 \equiv [y_1^1,y_2^1]$, divided by the total cost $C^t(y^0,w)$ of producing the vector of observed period 0 procedure outputs, $y^0 \equiv [y_1^0,y_2^0]$, where in both cases, we use the technology of period t and assume that the service providers face the vector of reference input prices, $w \equiv [w_1,w_2]$, where w_i is a reference vector of input prices for sector i. Thus for each choice of technology (i.e., t could equal 0 or 1) and for each choice of a reference vector of input prices w, we obtain a (different) cost based output quantity index.

Following the example of Konüs (1939), it is natural to single out two special cases of the family of output quantity indexes defined by (A10): one choice where we use the period 0 technology and set the reference prices equal to the period 0 input prices $w^0 \equiv [w_1^0,w_2^0]$

24 Our approach is a reasonably straightforward adaptation of the earlier work on theoretical price and quantity indexes by Konüs (1939), Fisher and Shell (1972), Samuelson and Swamy (1974), Archibald (1977) and Diewert (1980; 461) (1983; 1054-1083).

and another choice where we use the period 1 technology and set the reference prices equal to the period 1 input prices $w^1 \equiv [w_1^1,w_2^1]$. Thus define these special cases as α_0 and α_1:

$$
\begin{aligned}
\text{(A11)}\quad \alpha_0 &\equiv \alpha(y^0,y^1,w^0,0) \\
&= [c_1^0(w_1^0)y_1^1 + c_2^0(w_2^0)y_2^1]/[c_1^0(w_1^0)y_1^0 + c_2^0(w_2^0)y_2^0] && \text{using (A10)} \\
&= [p_1^0y_1^1 + p_2^0y_2^1]/[p_1^0y_1^0 + p_2^0y_2^0] && \text{using (A6)} \\
&= Q_L && \text{using (10);}
\end{aligned}
$$

$$
\begin{aligned}
\text{(A12)}\quad \alpha_1 &\equiv \alpha(y^0,y^1,w^1,1) \\
&= [c_1^1(w_1^1)y_1^1 + c_2^1(w_2^1)y_2^1]/[c_1^1(w_1^1)y_1^0 + c_2^1(w_2^1)y_2^0] && \text{using (A10)} \\
&= [p_1^1y_1^1 + p_2^1y_2^1]/[p_1^1y_1^0 + p_2^1y_2^0] && \text{using (A6)} \\
&= Q_P && \text{using (11).}
\end{aligned}
$$

Thus the theoretical cost based output quantity index α_0 that uses the period 0 technology and period 0 input prices w^0 is equal to the observable Laspeyres output quantity index Q_L that was defined earlier in the main text by (10) and the theoretical cost based output quantity index α_1 that uses the period 1 technology and period 1 input prices w^1 is equal to the observable Paasche output quantity index Q_P that was defined earlier by (11). Since both theoretical output quantity indexes, α_0 and α_1, are equally representative, a single estimate of cost based output quantity growth should be set equal to a symmetric average of these two estimates. We will choose the geometric mean as our preferred symmetric average[25] and thus our preferred theoretical measure of cost based output quantity growth is the following Fisher type theoretical index, α_F:

$$
\begin{aligned}
\text{(A13)}\quad \alpha_F &\equiv [\alpha_0\alpha_1]^{1/2} \\
&= [Q_LQ_P]^{1/2} && \text{using (A11) and (A12)} \\
&= Q_F && \text{using definition (12).}
\end{aligned}
$$

Thus our preferred measure of cost based output growth is equal to the observable Fisher quantity index, Q_F. which was defined earlier by (12) in the main text.

We now turn our attention to theoretical measures of input price growth. We now use the total cost function in order to define a *family of input price indexes*, $\beta(w^0,w^1,y,t)$, as follows:

$$
\begin{aligned}
\text{(A14)}\quad \beta(w^0,w^1,y,t) &\equiv C^t(y,w^1)/C^t(y,w^0) \\
&= [c_1^t(w_1^1)y_1 + c_2^t(w_2^1)y_2]/[c_1^t(w_1^0)y_1 + c_2^t(w_2^0)y_2] && \text{using (A8).}
\end{aligned}
$$

Thus the theoretical output quantity index $\beta(w^0,w^1,y,t)$ defined by (A14) is equal to the (hypothetical) total cost $C^t(y,w^1)$ of producing the reference vector of outputs, $y \equiv [y_1,y_2]$, when the service providers face the period 1 observed vector of input prices w^1, divided by the total cost $C^t(y,w^0)$ of producing the same reference vector of outputs, y, when the service providers face the period 0 observed vector of input prices w^0, where in both

[25] Diewert (1997) explained why the geometric mean is a good choice for the symmetric average.

cases, we use the technology of period t. Thus for each choice of technology (i.e., t could equal 0 or 1) and for each choice of a reference vector of output quantities y, we obtain a (different) input price index.

Again following the example of Konüs (1939) in his analysis of the true cost of living index, it is natural to single out two special cases of the family of input price indexes defined by (A14): one choice where we use the period 0 technology and set the reference quantities equal to the period 0 quantities $y^0 \equiv [y_1^0,y_2^0]$ and another choice where we use the period 1 technology and set the reference quantities equal to the period 1 quantities $y^1 \equiv [y_1^1,y_2^1]$. Thus define these special cases as β_0 and β_1:

$$\begin{aligned} \text{(A15)}\ \beta_0 &\equiv \beta(w^0,w^1,y^0,0) \\ &= [c_1^0(w_1^1)y_1^0 + c_2^0(w_2^1)y_2^0]/[c_1^0(w_1^0)y_1^0 + c_2^0(w_2^0)y_2^0] && \text{using (A14)} \\ &= [c_1^0(w_1^1)y_1^0 + c_2^0(w_2^1)y_2^0]/[w_1^0\cdot x_1^0 + w_2^0\cdot x_2^0] && \text{using (A9);} \end{aligned}$$

$$\begin{aligned} \text{(A16)}\ \beta_1 &\equiv \beta(w^0,w^1,y^1,1) \\ &= [c_1^1(w_1^1)y_1^1 + c_2^1(w_2^1)y_2^1]/[c_1^1(w_1^0)y_1^1 + c_2^1(w_2^0)y_2^1] && \text{using (A14)} \\ &= [w_1^1\cdot x_1^1 + w_2^1\cdot x_2^1]/[c_1^1(w_1^0)y_1^1 + c_2^1(w_2^0)y_2^1] && \text{using (A9).} \end{aligned}$$

We now encounter a problem: the hypothetical unit costs $c_i^0(w_i^1)$ and $c_i^1(w_i^0)$ which appear in (A15) and (A16) are not observable so we cannot calculate the theoretical input price indexes β_0 and β_1. However, we can find bounds to these indexes as well as first order Taylor series approximations to them, which are observable, as we now show.

As mentioned above, the unit cost functions, $c_i^t(w_i)$ are concave functions in their input price variables w_i. It is well known that the first order Taylor series approximation to a concave function lies above (or is coincident with) the concave function[26] so we have the following inequalities:

$$\begin{aligned} \text{(A17)}\ c_i^0(w_i^1) &\leq c_i^0(w_i^0) + \nabla_w c_i^0(w_i^0)\cdot[w_i^1 - w_i^0]\ ; && i = 1,2 \\ &= w_i^1\cdot\nabla_w c_i^0(w_i^0) && \text{since } w_i^1\cdot\nabla_w c_i^0(w_i^0) = c_i^0(w_i^0)^{27} \\ &= w_i^1\cdot a_i^0 && \text{using (A5).} \end{aligned}$$

The gap between the right and left hand sides of (A17) represents *input substitution bias*. In the main text, we assumed Leontief no substitution type procedure production functions and so there was no substitution bias; i.e., under our main text assumptions, the inequalities in (A17) were equalities. Now multiply both sides of inequality i in (A17) by y_i^0 and we obtain the following inequalities:

$$\begin{aligned} \text{(A18)}\ c_i^0(w_i^1)y_i^0 &\leq w_i^1\cdot a_i^0 y_i^0\ ; && i = 1,2 \\ &= w_i^1\cdot x_i^0 && \text{using (A4).} \end{aligned}$$

[26] See Fenchel (1953) or Mangasarian (1969; 84).
[27] This follows from Euler's Theorem on homogeneous functions and the fact that $c_i^0(w_i)$ is linearly homogeneous in the components of the input price vector w_i.

In a similar fashion, we can establish the following inequalities:

$$\text{(A19)}\quad c_i^1(w_i^0)y_i^1 \le w_i^0 \cdot x_i^1\,; \qquad\qquad i = 1,2.$$

We now return to the theoretical input price indexes defined by (A15) and (A16). From (A15), we have:

$$\begin{aligned}\text{(A20)}\quad \beta_0 &= [c_1^0(w_1^1)y_1^0 + c_2^0(w_2^1)y_2^0]/[w_1^0 \cdot x_1^0 + w_2^0 \cdot x_2^0] \\ &\le [w_1^1 \cdot x_1^0 + w_2^1 \cdot x_2^0]/[w_1^0 \cdot x_1^0 + w_2^0 \cdot x_2^0] \qquad \text{using (A18)} \\ &\equiv P_L^*\end{aligned}$$

where the observable Laspeyres input price index P_L^* is defined as $[w_1^1 \cdot x_1^0 + w_2^1 \cdot x_2^0]/[w_1^0 \cdot x_1^0 + w_2^0 \cdot x_2^0]$. Similarly, from (A16), we have:

$$\begin{aligned}\text{(A21)}\quad \beta_1 &= [w_1^1 \cdot x_1^1 + w_2^1 \cdot x_2^1]/[c_1^1(w_1^0)y_1^1 + c_2^1(w_2^0)y_2^1] \\ &\ge [w_1^1 \cdot x_1^1 + w_2^1 \cdot x_2^1]/[w_1^0 \cdot x_1^1 + w_2^0 \cdot x_2^1] \qquad \text{using (A19)} \\ &\equiv P_P^*\end{aligned}$$

where the observable Paasche input price index P_P^* is defined as $[w_1^1 \cdot x_1^1 + w_2^1 \cdot x_2^1]/[w_1^0 \cdot x_1^1 + w_2^0 \cdot x_2^1]$. Thus the theoretical input price index β_0 is bounded from above by the observable Laspeyres input price index P_L^* and the theoretical input price index β_1 is bounded from below by the observable Paasche input price index P_P^*. In both cases, the gap between the theoretical index and the observable index is due to input substitution bias, which goes in opposite directions.

Looking at the first line in (A17), it can be seen that the right hand sides of (A18) and (A19) are first order Taylor series approximations to the corresponding left hand side entries. This means that P_L^* is a first order approximation to the theoretical input price index β_0 and P_P^* is a first order approximation to the theoretical input price index β_1.

Since both theoretical input price indexes, β_0 and β_1, are equally representative, a single estimate of input price change should be set equal to a symmetric average of these two estimates. We again choose the geometric mean as our preferred symmetric average and thus our preferred theoretical measure of input price growth is the following Fisher type theoretical index, β_F:

$$\begin{aligned}\text{(A22)}\quad \beta_F &\equiv [\beta_0\beta_1]^{1/2} \\ &\approx [P_L^* P_P^*]^{1/2} \\ &\equiv P_F^*\end{aligned}$$

where the Fisher (1922) index of input price change, P_F^*, is defined as the geometric mean of the Laspeyres and Paasche input price indexes. Given the fact that P_L^* is a first order approximation to β_0 and P_P^* is a first order approximation to β_1, it is obvious that P_F^* is at least a first order approximation to the theoretical input price index β_F. But in most cases, the approximation of P_F^* to β_F will be much better than a first order approximation since the upward bias in P_L^* will generally offset the downward bias in P_P^*.

We now define our last family of theoretical indexes. We again use the total cost function in order to define a *family of reciprocal indexes of technical progress*, $\gamma(y,w)$, as follows:

$$\text{(A23)}\quad \gamma(y,w) \equiv C^1(y,w)/C^0(y,w) = [c_1^1(w_1)y_1 + c_2^1(w_2)y_2]/[c_1^0(w_1)y_1 + c_2^0(w_2)y_2] \qquad \text{using (A8).}$$

The theoretical reciprocal technical progress index $\gamma(y,w)$ defined by (A23) is equal to the (hypothetical) total cost $C^1(y,w)$ of producing the reference vector of outputs, $y \equiv [y_1,y_2]$, when the service providers face the reference vector of input prices w using the period 1 technology, divided by the total cost $C^0(y,w)$ of producing the same reference vector of outputs, y, and facing the same reference vector of input prices w, where we now use the period 0 technology.[28] Thus $\gamma(y,w)$ is a measure of the *proportional reduction in costs* that occurs due to technical progress between periods 0 and 1 and it can be seen that this is an inverse measure of technical progress. For each choice of a reference vector of output quantities y and reference vector of input prices w, we obtain a (different) measure of exogenous cost reduction.

Instead of singling out the reference vectors y and w that appear in the definition of $\gamma(y,w)$ to be the period t quantity and price vectors (y^t,w^t) for t = 0,1, we will choose the mixed vectors (y^0,w^1) and (y^1,w^0) for special attention. The reason for these rather odd looking choices will be explained below.

We want to explain the growth in total costs going from period 0 to 1, $C^1(y^1,w^1)/C^0(y^0,w^0)$, as the product of 3 growth factors:

- Growth in outputs; i.e., a factor of the form $\alpha(y^0,y^1,w,t)$ defined above by (A10);
- Growth in input prices; i.e., a factor of the form $\beta(w^0,w^1,y,t)$ defined by (A14) and
- Exogenous reduction in costs due to technical progress; i.e., a factor of the form $\gamma(y,w)$ defined by (A23).

Simple algebra shows that we have the following decompositions of the cost ratio $C^1(y^1,w^1)/C^0(y^0,w^0)$ into explanatory factors of the above type:[29]

$$\begin{aligned}\text{(A24)}\quad C^1(y^1,w^1)/C^0(y^0,w^0) &= [C^1(y^1,w^1)/C^1(y^0,w^1)][C^0(y^0,w^1)/C^0(y^0,w^0)][C^1(y^0,w^1)/C^0(y^0,w^1)] \\ &= \alpha_1\beta_0\gamma(y^0,w^1) \qquad \text{using definitions (A12), (A15) and (A23);}\end{aligned}$$

[28] This is a cost function analogue to the revenue function definitions of technical progress defined by Diewert (1983; 1063-1064), Diewert and Morrison (1986) and Kohli (1990).

[29] The decompositions of cost growth given by (A24) and (A25) are nonparametric analogues to the parametric revenue growth decompositions obtained by Diewert and Morrison (1986), Kohli (1990), (1991) (2003) and Fox and Kohli (1998) into explanatory factors.

$$(A25)\ C^1(y^1,w^1)/C^0(y^0,w^0)$$
$$= [C^0(y^1,w^0)/C^0(y^0,w^0)][C^1(y^1,w^1)/C^1(y^1,w^0)][C^1(y^1,w^0)/C^0(y^1,w^0)]$$
$$= \alpha_0\beta_1\gamma(y^1,w^0) \qquad \text{using definitions (A11), (A16) and (A23).}$$

The above decompositions show that the two special cases of $\gamma(y,w)$ defined by (A23) of particular interest are defined by (A26) and (A27) below:

$$(A26)\ \gamma(y^0,w^1) \equiv C^1(y^0,w^1)/C^0(y^0,w^1) \qquad \text{using (A23)}$$
$$= [c_1{}^1(w_1{}^1)y_1{}^0 + c_2{}^1(w_2{}^1)y_2{}^0]/[c_1{}^0(w_1{}^1)y_1{}^0 + c_2{}^0(w_2{}^1)y_2{}^0] \qquad \text{using (A8);}$$

$$(A27)\ \gamma(y^1,w^0) \equiv C^1(y^1,w^0)/C^0(y^1,w^0) \qquad \text{using (A23)}$$
$$= [c_1{}^1(w_1{}^0)y_1{}^1 + c_2{}^1(w_2{}^0)y_2{}^1]/[c_1{}^0(w_1{}^0)y_1{}^1 + c_2{}^0(w_2{}^0)y_2{}^1] \qquad \text{using (A8).}$$

We will now work out observable first order approximations (and observable bounds) to the two specific measures of reciprocal technical progress defined by (A26) and (A27). From the second equation in (A26), we have:

$$(A28)\ C^1(y^0,w^1) = c_1{}^1(w_1{}^1)y_1{}^0 + c_2{}^1(w_2{}^1)y_2{}^0$$
$$= p_1{}^1y_1{}^0 + p_2{}^1y_2{}^0 \qquad \text{using (A6)}$$
$$= p^1\cdot y^0.$$

Since $C^0(y^0,w)$ is concave in the components of the input price vector w, $C^0(y^0,w)$ regarded as a function of w will be bounded from above by its first order Taylor series approximation around the point w^0 so the following inequality will be satisfied:

$$(A29)\ C^0(y^0,w^1) \leq C^0(y^0,w^0) + \nabla_w C^0(y^0,w^0)\cdot[w^1 - w^0]$$
$$= w^1\cdot\nabla_w C^0(y^0,w^0) \qquad \text{since } w^0\cdot\nabla_w C^0(y^0,w^0) = C^0(y^0,w^0)^{30}$$
$$= w_1{}^1\cdot x_1{}^0 + w_2{}^1\cdot x_2{}^0 \qquad \text{using (A3) and (A8)}$$
$$\equiv w^1\cdot x^0.$$

Thus the period 0 total cost function $C^0(y^0,w^1)$, evaluated at the vector of period 0 observed outputs y^0 and the period 1 observed input prices w^1, is bounded from above by the inner product of the period 1 input price vector w^1 and the observed vector of inputs for period 0, x^0.[31] Note also from the first line of (A29) that $w^1\cdot x^0$ is a first order Taylor series approximation to the unobserved cost $C^0(y^0,w^1)$. The results (A28) and (A29) can now be used in order to form a bound (and a first order approximation) to the measure of reciprocal technical progress $\gamma(y^0,w^1)$ defined by (A26):

$$(A30)\ \gamma(y^0,w^1) = C^1(y^0,w^1)/C^0(y^0,w^1)$$
$$\geq p^1\cdot y^0/w^1\cdot x^0 \qquad \text{using (A28) and (A29)}$$

---

[30] This follows from Euler's Theorem on homogeneous functions and the fact that $C^0(y^0,w)$ is linearly homogeneous in the components of the input price vector w.

[31] This result can also be established by noting that x^0 is a feasible (but not necessarily optimal) solution to the industry cost minimization problem defined by $C^0(y^0,w^1)$.

$$= [p^1 \cdot y^0/p^1 \cdot y^1]/[w^1 \cdot x^0/w^1 \cdot x^1] \qquad \text{using (A7) for } t = 1$$
$$= [w^1 \cdot x^1/w^1 \cdot x^0]/[p^1 \cdot y^1/p^1 \cdot y^0] \qquad \text{rearranging terms}$$
$$= Q_P^*/Q_P \qquad \text{using (7) and (11)}$$
$$= [Q_P/Q_P^*]^{-1}$$

where Q_P is the Paasche output quantity index defined by (11) and Q_P^* is the Paasche input quantity index defined by (7) in the main text. Note that Q_P divided by Q_P^* is the *Paasche productivity index*. Thus (A30) tells us that the theoretical measure of reciprocal technical progress defined by (A26), $\gamma(y^0,w^1)$, is bounded from above by the reciprocal of the observable Paasche productivity index, Q_P/Q_P^*. Moreover, it can be seen that $[Q_P/Q_P^*]^{-1}$ is also a first order approximation to the theoretical index $\gamma(y^0,w^1)$.

The above algebra can be repeated with minor modifications in order to derive a bound for the theoretical index $\gamma(y^1,w^0)$ defined by (A27). Thus we have:

$$\text{(A31)} \quad C^0(y^1,w^0) = c_1{}^0(w_1{}^0)y_1{}^1 + c_2{}^0(w_2{}^0)y_2{}^1$$
$$= p_1{}^0 y_1{}^1 + p_2{}^0 y_2{}^1 \qquad \text{using (A6)}$$
$$= p^0 \cdot y^1.$$

Since $C^1(y^1,w)$ is concave in the components of the input price vector w, $C^1(y^1,w)$ regarded as a function of w will be bounded from above by its first order Taylor series approximation around the point w^1 so the following inequality will be satisfied:

$$\text{(A32)} \quad C^1(y^1,w^0) \le C^1(y^1,w^1) + \nabla_w C^1(y^1,w^1) \cdot [w^0 - w^1]$$
$$= w^0 \cdot \nabla_w C^1(y^1,w^1) \qquad \text{since } w^1 \cdot \nabla_w C^1(y^1,w^1) = C^1(y^1,w^1)$$
$$= w_1{}^0 \cdot x_1{}^1 + w_2{}^0 \cdot x_2{}^1 \qquad \text{using (A3) and (A8)}$$
$$\equiv w^0 \cdot x^1.$$

Thus the period 1 total cost function $C^1(y^1,w^0)$, evaluated at the vector of period 1 observed outputs y^1 and the period 0 observed input prices w^0, is bounded from above by the inner product of the period 0 input price vector w^0 and the observed vector of inputs for period 1, x^1.[32] Note also from the first line of (A32) that $w^0 \cdot x^1$ is a first order Taylor series approximation to the unobserved cost $C^1(y^1,w^0)$. The results (A31) and (A32) can now be used in order to form a bound (and a first order approximation) to the measure of reciprocal technical progress $\gamma(y^1,w^0)$ defined by (A27):

$$\text{(A33)} \quad \gamma(y^1,w^0) = C^1(y^1,w^0)/C^0(y^1,w^0)$$
$$\le w^0 \cdot x^1/p^0 \cdot y^1 \qquad \text{using (A31) and (A32)}$$
$$= [w^0 \cdot x^1/w^0 \cdot x^0]/[p^0 \cdot y^1/p^0 \cdot y^0] \qquad \text{using (A7) for } t = 0$$
$$= Q_L^*/Q_L \qquad \text{using (6) and (10)}$$
$$= [Q_L/Q_L^*]^{-1}$$

[32] This result can also be established by noting that x^1 is a feasible (but not necessarily optimal) solution to the industry cost minimization problem defined by $C^1(y^1,w^0)$.

where Q_L is the Laspeyres output quantity index defined by (10) and Q_L^* is the Laspeyres input quantity index defined by (6) in the main text. Note that Q_L divided by Q_L^* is the *Laspeyres productivity index*. Thus (A33) tells us that the theoretical measure of reciprocal technical progress defined by (A27), $\gamma(y^1,w^0)$, is bounded from above by the reciprocal of the observable Laspeyres productivity index, Q_L/Q_L^*. Moreover, it can be seen that $[Q_L/Q_L^*]^{-1}$ is also a first order approximation to the theoretical index $\gamma(y^1,w^0)$.

Since the two cost decompositions for the rate of growth of cost, $C^1(y^1,w^1)/C^0(y^0,w^0)$, given by (A24) and (A25) are equally valid, we will take the geometric average of these two decompositions to obtain our preferred overall cost decomposition. This leads to the following theoretical decomposition of $C^1(y^1,w^1)/C^0(y^0,w^0)$ into explanatory factors:

$$\text{(A34)}\quad C^1(y^1,w^1)/C^0(y^0,w^0) = \alpha_F\,\beta_F\,\gamma_F$$

where the Fisher type output quantity growth factor α_F is defined by (A13), the Fisher type input price growth factor β_F is defined by (A22) and the *Fisher type reciprocal measure of technical progress* γ_F is defined as follows:

$$\text{(A35)}\quad \gamma_F \equiv [\gamma(y^0,w^1)\gamma(y^1,w^0)]^{1/2}.$$

Using our first order approximations given by (A30) and (A33), it can be seen that an observable first order approximation to γ_F is the reciprocal of the *Fisher productivity index* Q_F/Q_F^*; i.e., we have:

$$\begin{aligned}
\text{(A36)}\quad \gamma_F &\approx \{[Q_P^*/Q_P][Q_L^*/Q_L]\}^{1/2} && \text{using (A30) and (A33)}\\
&= [Q_F^*/Q_F] && \text{using (8) and (12)}\\
&= [Q_F/Q_F^*]^{-1}.
\end{aligned}$$

However, since the substitution biases in the first order approximations given by (A30) and (A33) go in opposite directions, the approximation to the theoretical index γ_F given by the right hand side of (A36) will generally be much closer than a first order approximation.

We now combine the theoretical cost decomposition defined by (A34) with the exact result (A13) and the approximate results (A22) and (A36):

$$\begin{aligned}
\text{(A37)}\quad C^1(y^1,w^1)/C^0(y^0,w^0) &= \alpha_F\,\beta_F\,\gamma_F && \text{using (A34)}\\
&\approx Q_F\,P_F^*\,[Q_F/Q_F^*]^{-1} && \text{using (A13), (A22) and (A36).}
\end{aligned}$$

Thus (one plus) the rate of growth of industry cost, $C^1(y^1,w^1)/C^0(y^0,w^0)$, is approximately equal to (one plus) the rate of output growth defined by the Fisher output quantity index, Q_F, times (one plus) the rate of growth of input prices defined by the Fisher input price index, P_F^*, times the reciprocal of (one plus) the Fisher rate of productivity growth, Q_F/Q_F^*. However, it turns out that the left hand side of (A37) is *identically equal* to the product of the explanatory factors on the right hand side of (A37), since it can be shown that the following identity holds:[33]

(A38) $[C^1(y^1,w^1)/C^0(y^0,w^0)]/P_F^* = Q_F^*$;

i.e., the Fisher implicit input quantity index, $[w^1 \cdot x^1/w^0 \cdot x^0]/P_F^*$, is equal to the direct Fisher input quantity index, Q_F^*.

Note that the above results are entirely nonparametric. Thus we have generalized (to a reasonable degree of approximation) the results derived in the main text under the assumption that the procedure production functions were of the no substitution variety to the case where the procedure production functions are general ones.

It is possible to further generalize our results from the case where the procedure functions are independent to the case where there are shared overheads between the procedures.[34] In this more general framework, the health sector's period t total cost function $C^t(y,w)$ is no longer defined as the sum of the two procedure cost functions, $C_1^t(y_1,w_1)$ plus $C_2^t(y_2,w_2)$ as in (A8), but is simply a general nonjoint cost function. The regularity conditions that we impose on each $C^t(y,w)$ is that it is a nonnegative, jointly continuous differentiable function in its variables (y,x) and that it is linearly homogeneous[35], nondecreasing and convex[36] in the components of y for fixed w and linearly homogeneous, nondecreasing and concave in the components of w for each fixed y. As usual, we assume cost minimizing behavior in periods t = 0,1 and that we can observe the period t industry output vector, $y^t \equiv [y_1^t,y_2^t]$, the period t aggregate input vector $x^t \equiv [x_1^t,x_2^t]$ and the corresponding vector of input prices $w^t \equiv [w_1^t,w_2^t]$ for t = 0,1. As usual, Shephard's Lemma tells us that the period t vector of inputs is equal to the vector of first order partial derivatives of the period t cost function with respect to the components of the input price vector; i.e., we have:

(A39) $x^t = \nabla_w C^t(y^t,w^t)$; t = 0,1.

The period t vector of *marginal cost output prices* $p^t \equiv [p_1^t,p_2^t]$ is defined as the vector of first order partial derivatives of the period t cost function with respect to the components of the output vector:

(A40) $p^t \equiv \nabla_y C^t(y^t,w^t)$; t = 0,1.

It should be noted that the linear homogeneity properties of $C^t(y,w)$ in y and w separately imply the following equalities:

(A41) $C^t(y^t,w^t) = w^t \cdot x^t = p^t \cdot y^t$; t = 0,1.

---

[33] See Fisher (1922).
[34] Yu (2008) uses this joint cost function framework to measure health and other nonmarket outputs.
[35] This assumption means that the overall technology is subject to constant returns to scale; i.e., the period t technology set S^t is a cone.
[36] This restriction means that the overall technology set S^t is a convex set.

We can now modify the above analysis in this appendix, letting the marginal cost prices p^t defined by (A40) replace our old unit cost prices.

In particular, we use the first line in (A10) in order to define a new family of cost based output quantity indexes as $\alpha(y^0,y^1,w,t) \equiv C^t(y^1,w)/C^t(y^0,w)$ and we again use the first line in (A11) and (A12), to define the two specific output quantity indexes α_0 as $\alpha(y^0,y^1,w^0,0)$ and α_1 as $\alpha(y^0,y^1,w^1,1)$. However, in our new more general model, we no longer obtain the equalities in (A11) and (A12); instead, we obtain the following first order approximations and bounds:

$$
\begin{aligned}
\text{(A42)}\ \alpha_0 &\equiv C^0(y^1,w^0)/C^0(y^0,w^0) \\
&= C^0(y^1,w^0)/p^0\cdot y^0 && \text{using (A41)} \\
&\geq \{C^0(y^0,w^0) + \nabla_y C^0(y^0,w^0)\cdot[y^1 - y^0]\}/p^0\cdot y^0 \\
& && \text{using the convexity of } C^0(y,w^0) \text{ in } y^{37} \\
&= \nabla_y C^0(y^0,w^0)\cdot y^1/p^0\cdot y^0 && \text{using } \nabla_y C^0(y^0,w^0)\cdot y^0 = C^0(y^0,w^0) \\
&= p^0\cdot y^1/p^0\cdot y^0 && \text{using (A40) for } t = 0 \\
&= Q_L && \text{using (10);}
\end{aligned}
$$

$$
\begin{aligned}
\text{(A43)}\ \alpha_1 &\equiv C^1(y^1,w^1)/C^1(y^0,w^1) \\
&= p^1\cdot y^1/C^1(y^0,w^1) && \text{using (A41)} \\
&\leq p^1\cdot y^1/\{C^1(y^1,w^1) + \nabla_y C^1(y^1,w^1)\cdot[y^0 - y^1]\} && \text{using the convexity of } C^1(y,w^1) \text{ in } y \\
&= p^1\cdot y^1/\{\nabla_y C^1(y^1,w^1)\cdot y^0\} && \text{using } \nabla_y C^1(y^1,w^1)\cdot y^1 = C^1(y^1,w^1) \\
&= p^1\cdot y^1/p^1\cdot y^0 && \text{using (A40) for } t = 1 \\
&= Q_P && \text{using definition (11).}
\end{aligned}
$$

Thus the ordinary Laspeyres output quantity index Q_L (using marginal cost prices) is no longer equal to the theoretical output index α_0 defined by the first line in (A42) but is only a first order approximation and a lower bound. Similarly, ordinary Paasche output quantity index Q_P (using marginal cost prices) is no longer equal to the theoretical output index α_1 defined by the first line in (A43) but is only a first order approximation and an upper bound to this theoretical output quantity index. However, as before, the Fisher theoretical output quantity index α_F defined as the geometric mean of α_0 and α_1 will be approximated by the Fisher output quantity index, $Q_F \equiv [Q_L Q_P]^{1/2}$, and due to the offsetting substitution biases in (A42) and (A43), Q_F will generally approximate α_F to an accuracy that is greater than a first order approximation.

The analysis associated with (A14)-(A22) goes through with obvious modifications using our more general model and so we will not repeat this analysis.

We can again define the family of reciprocal indexes of technical progress using the first line in (A23) as $\gamma(y,w) \equiv C^1(y,w)/C^0(y,w)$. As before, the factorizations of cost growth given by (A24) and (A25) continue to be valid in this more general framework and so we

[37] The first order Taylor series approximation to a convex function lies below (or is coincident with) the function.

need empirically observable approximations to the two indexes of technical progress defined by (A26) and (A27). The inequalities in (A29) and (A32) continue to be valid but the equalities in (A28) and (A31) are no longer valid in our more general model and need to be replaced by the following inequalities which were derived in (A43) and (A42):

$$\text{(A44)}\quad C^1(y^0,w^1) \geq C^1(y^1,w^1) + \nabla_y C^1(y^1,w^1)\cdot[y^0 - y^1]\}$$
$$= p^1\cdot y^0 \,;$$

$$\text{(A45)}\quad C^0(y^1,w^0) \geq C^0(y^0,w^0) + \nabla_y C^0(y^0,w^0)\cdot[y^1 - y^0]$$
$$= p^0\cdot y^1.$$

Using definition (A26) and the inequalities (A29) and (A44) establishes the following inequality:

$$\begin{aligned}
\text{(A46)}\quad \gamma(y^0,w^1) &= C^1(y^0,w^1)/C^0(y^0,w^1) && \text{using definition (A26)}\\
&\geq p^1\cdot y^0/w^1\cdot x^0 && \text{using (A44) and (A29)}\\
&= [p^1\cdot y^0/p^1\cdot y^1]/[w^1\cdot x^0/w^1\cdot x^1] && \text{using (A41) for } t = 1\\
&= Q_P{}^*/Q_P && \text{using (7) and (11)}\\
&= [Q_P/Q_P{}^*]^{-1}.
\end{aligned}$$

Similarly, using definition (A27) and the inequalities (A29) and (A44) establishes the following inequality:

$$\begin{aligned}
\text{(A47)}\quad \gamma(y^1,w^0) &\equiv C^1(y^1,w^0)/C^0(y^1,w^0) && \text{using definition (A27)}\\
&\leq w^0\cdot x^1/p^0\cdot y^1 && \text{using (A32) and (A45)}\\
&= [w^0\cdot x^1/w^0\cdot x^0]/[p^0\cdot y^1/p^0\cdot y^0] && \text{using (A41) for } t = 0\\
&= Q_L{}^*/Q_L && \text{using (6) and (10)}\\
&= [Q_L/Q_L{}^*]^{-1}.
\end{aligned}$$

As before, define the Fisher type reciprocal measure of technical progress γ_F by (A35). The rest of the above analysis goes through with minor modifications. In particular, we still obtain the cost decomposition (A37) as the following exact equality:[38]

$$\begin{aligned}
\text{(A48)}\quad C^1(y^1,w^1)/C^0(y^0,w^0) &= \alpha_F\,\beta_F\,\gamma_F\\
&= Q_F\,P_F{}^*\,[Q_F/Q_F{}^*]^{-1}.
\end{aligned}$$

Note that it is *risky* to use either the Laspeyres or Paasche measures of productivity growth to approximate the corresponding theoretically correct measure of productivity growth since there are *two doses of substitution bias* between the left hand and right hand sides of (A46) and (A47); i.e., *the output and input substitution biases augment each other* when we use Paasche or Laspeyres productivity indexes instead of offsetting each

---

[38] However, in general, Q_F will not be exactly equal to its theoretical counterpart α_F and $P_F{}^*$ will not be exactly equal to its theoretical counterpart β_F. But when the three empirical growth factors are multiplied together, the various approximation errors cancel out to give an overall exact decomposition.

other. Thus whenever possible, we recommend the use of Fisher productivity indexes rather than the use of their Paasche or Laspeyres counterparts.

The reader may well wonder why we did not proceed directly to our final most general model of production instead of doing the additive cost model defined by (A8) where no joint costs were present. The problem is that *our most general model requires estimates of marginal cost prices* in order to implement the practical approximations to the theoretical indexes. Unfortunately, econometric estimation of joint cost functions will generally be required in order to estimate these marginal cost prices and econometric estimation of joint cost functions with general technical progress and the use of flexible functional forms is a generally hazardous exercise!

References

Archibald, R.B. (1977), "On the Theory of Industrial Price Measurement: Output Price Indexes", *Annals of Economic and Social Measurement* 6, 57-62.

Atkinson, Tony (2005), *Atkinson Review: Final Report; Measurement of Government Output and Productivity for the National Accounts*, New York: Palgrave Macmillan.

Diewert, W.E., (1974), "Applications of Duality Theory," pp. 106-171 in M.D. Intriligator and D.A. Kendrick (ed.), *Frontiers of Quantitative Economics*, Vol. II, Amsterdam: North-Holland. http://www.econ.ubc.ca/diewert/theory.pdf

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, D. Usher (ed.), Chicago: The University of Chicago Press.

Diewert, W.E. (1983), "The Theory of the Output Price Index and the Measurement of Real Output Change", pp. 1049-1113 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Diewert, W.E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", Discussion Paper No. 95-01, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W.E. (1997), "Commentary on Mathew D. Shapiro and David W. Wilcox: Alternative Strategies for Aggregating Price in the CPI", *The Federal Reserve Bank of St. Louis Review*, 79:3, 127-137.

Diewert, W.E. (1998), "Index Number Issues in the Consumer Price Index", *The Journal of Economic Perspectives*, 12:1, 47-58.

Diewert, W.E. and C.J. Morrison (1986), "Adjusting Output and Productivity Indexes for Changes in the Terms of Trade", *The Economic Journal* 96, 659-679.

Eurostat, IMF, OECD, UN and the World Bank (1993), *System of National Accounts 1993*, New York: The United Nations.

Fenchel, W. (1953), "Convex Cones, Sets and Functions", Lecture Notes at Princeton University, Department of Mathematics, Princeton, N.J.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Fisher, F.M. and K. Shell (1972), "The Pure Theory of the National Output Deflator", pp. 49-113 in *The Economic Theory of Price Indexes*, New York: Academic Press.

Fox, K.J. and U. Kohli (1998), "GDP Growth, Terms of Trade Effects and Total Factor Productivity", *Journal of International Trade and Economic Development* 7, 87-110.

Hicks, John R. (1940), "On the Valuation of the Social Income", *Economica* 7, 105-124.

International Labour Organization (ILO), Eurostat, IMF, OECD, World Bank and the UN (2004), *Consumer Price Index Manual: Theory and Practice*, Peter Hill (ed.), International Labour Organization, Geneva. http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm

Kohli, U. (1990), "Growth Accounting in the Open Economy: Parametric and Nonparametric Estimates", *Journal of Economic and Social Measurement* 16, 125-136.

Kohli, U. (1991), *Technology, Duality and Foreign Trade: The GNP Function Approach to Modeling Imports and Exports*, Ann Arbor: University of Michigan Press.

Kohli, U. (2003), "Growth Accounting in the Open Economy: International Comparisons", *International Review of Economics and Finance* 12, 417-435.

Konüs, A.A. (1939), "The Problem of the True Index of the Cost of Living", *Econometrica* 7, 10-29.

Lawrence, D. and W.E. Diewert (2006), "Regulating Electricity Networks: The ABC of Setting X in New Zealand", pp. 207-241 in *Performance Measurement and Regulation of Network Utilities*, T. Coelli and D. Lawrence (eds.), Cheltenham: Edward Elgar Publishing.

Mai, N. (2004), "Measuring Health Care Output in the UK: A Diagnosis Based Approach", *Economic Trends* 610 (September), Office of National Statistics UK, 64-73.

Mangasarian, O. (1969), *Nonlinear Programming*, New York: McGraw-Hill.

Pritchard, A. (2004), "Measuring Government Health Services Output in the UK National Accounts: The New Methodology and Further Analysis", *Economic Trends* 613 (December), Office of National Statistics UK, 69-81.

Samuelson, P.A. and S. Swamy (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis", *American Economic Review* 64, 566-593.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton N.J.: Princeton University Press.

Yu, K. (2008), "Measurement of Government Output", forthcoming in *Essays on Price and Productivity Measurement*, Volume 3, W.E. Diewert, B. Balk, D. Fixler, K. Fox, and A.O. Nakamura (eds.), Victoria: Trafford Publishing.

Yu, K. and R. Ariste (2008), "Comparisons of Hospital Output in Canada: National and International Perspectives", unpublished paper, Department of Economics, Lakehead University.