Decompositions of Profitability Change using Cost Functions

W. Erwin Diewert[1]
Revised May 7, 2012.
Discussion Paper 10-07,
Department of Economics,
University of British Columbia,
Vancouver, B.C.,
Canada, V6T 1Z1.
Email: diewert@econ.ubc.ca

Abstract

The paper presents a decomposition of a production unit's cost ratio over two periods into explanatory factors. The explanatory factors are growth in the unit's cost efficiency, output growth, changes in input prices and technical progress. In order to implement the decomposition, an estimate of the industry's best practice cost function for the two periods under consideration is required. Profitability at a period of time is defined as the value of outputs produced by a production unit divided by the corresponding cost. Using the earlier work by Balk and O'Donnell, the paper provides a decomposition of profitability growth over two periods into various explanatory factors that are similar to the cost ratio decomposition except that change in outputs explanatory factor is replaced by two separate factors: an index of output price growth and a measure of returns to scale.

Keywords

Measurement of output, input and productivity, nonmarket sector, cost functions, duality theory, marginal cost prices, technical progress, returns to scale, total factor productivity, profitability, cost efficiency, Fisher ideal indexes, margin growth, best practice cost functions.

JEL Classification Numbers

C43, D24, D61, E23, H44, O47.

1. Introduction

In this paper, we will adapt some of the ideas about productivity and profitability decompositions that were explained in Diewert (2011) and O'Donnell (2008) (2010).

[1] Department of Economics, the University of British Columbia, Vancouver, Canada, V6T and School of Economics, the University of New South Wales. The author thanks Bert Balk, Chris O'Donnell, Antonio Peyrache, Prasada Rao, William Schworm, Valentin Zelenyuk and some unknown referees for helpful discussions and comments and the Social Science and Humanities Research Council of Canada, the Center for Efficiency and Productivity Analysis at the University of Queensland, Brisbane and the Center for Applied Economic Research at the University of New South Wales for financial support.

O'Donnell in his work provided some very useful decompositions that depended on the existence of various output and input quantity aggregates. In the present paper, we suggest that if best practice cost functions are available, then it is "reasonable" to use these best practice cost functions to form aggregates even though some aggregates do not necessarily satisfy all of the axioms that O'Donnell regards as being desirable axioms for a quantity index.[2]

Basically, we will look at the productivity performance of a production unit over two time periods. We assume that we have knowledge of a "best practice" cost function for the production unit for each of the two time periods. This best practice cost function could be the result of a Data Envelopment Analysis exercise or it could arise from an econometric study of the production units in a suitable peer group.[3]

In section 2, we will use these best practice cost functions to decompose a production unit's cost growth into explanatory factors which are: efficiency growth, changes in output quantities, input price growth and technical progress.

In section 3, we assume that output prices are available for the production unit in each period and we use these output prices in order to form output aggregates. In this section, our focus is on obtaining a decomposition of profitability growth into explanatory factors, where profitability is the ratio of the value of outputs produced in a period to the total cost of producing those outputs. We draw on the cost growth decomposition developed in section 2 and our main result is the profitability decomposition given by (34). We show how this decomposition is related to similar decompositions that exist in the literature.

Section 4 explores briefly the problem of finding maximum Total Factor Productivity (TFP) output combinations in each period. These TFP maximizing output combinations also maximize profitability for a cost efficient firm.

Section 5 concludes.

2. Decompositions of Cost Growth into Explanatory Factors

Before we look at changes in *profitability* (this is the ratio of the value of output in a period to the corresponding value of input)[4] for a production unit (establishment, firm, industry, economy), we will first attempt to obtain a decomposition of the production unit's *cost ratio* into explanatory factors. The decompositions that we develop in this section will prove to be useful in the following section.

---

[2] See also the discussion in Balk (1998; 90) on desirable axioms for a quantity index in the context of production theory.

[3] Alternatively, the reader can simply assume that the production unit is cost efficient in each period and we have an estimated cost function for the unit for each period. In this case, the inequalities in equations (3) and (4) become equalities and the cost efficiency index $\varepsilon(e^0,e^1)$ defined by (5) below is set equal to unity.

[4] Balk (2003; 9-10) introduced this terminology.

Before starting our formal analysis, it is necessary to justify why it is useful to work out a model of production unit efficiency in the context of a cost function model. Many countries have huge public sectors where production units provide goods and services to the public either free of charge or at prices that do not reflect either their costs of production or their desirability to purchasers. However, these public enterprises all hire labour, purchase intermediate inputs and rent (or own) capital equipment and structures. Moreover, information on the prices and quantities of these inputs is generally available in the government or enterprise accounts. If in addition, information on the quantities of outputs produced by these public enterprises is available, then we are in a position to look at the cost efficiency of these production units relative to a peer group of similar units. Under these conditions, it is possible for a productivity analyst to construct a *best practice cost function* that gives the minimum cost of producing the vector of outputs actually produced in each time period by each production unit in the peer group using the most efficient technology that is available to the peer group and using the input prices that the unit faces in the period under consideration. Thus the unit's actual cost in the period can be compared to the corresponding best practice cost (this is the unit's cost efficiency) and we can look at how the unit's cost efficiency evolves over time and decompose this evolution into explanatory factors. This analysis can take place without the analyst knowing output prices. This is the informational setup that will be utilized in the present section. In the next section, we will add the assumption that meaningful output prices are also available to the analyst. However, in the next section, the best practice cost function will continue to play a prominent role in the analysis of the production unit's efficiency over time.[5]

We assume that we can observe the strictly positive *period t input price vector*, $w^t >> 0_N$, the corresponding N dimensional nonnegative, nonzero *input quantity vector* $x^t > 0_N$ [6] for a particular production unit for periods t = 0,1. The *observed cost*, c^t, for the production unit under consideration for period t is:

$$(1)\ c^t \equiv w^t \cdot x^t ; \qquad t = 0,1.$$

We also assume that we can observe the M dimensional nonnegative, nonzero output quantity vector produced by the unit during period t, $y^t > 0_M$ for t = 0,1.

Our next assumption is much stronger than the above assumptions, which are not at all restrictive. We now assume that for each period t, there exists a *best practice technology* that the particular production unit under consideration could potentially access. Thus for each period t, there exists a best practice technology set, S^t, that defines a set of feasible

[5] Our contention is that virtually every nonmarket production unit will face market prices for at least some inputs and so it is reasonable to ask the unit to minimize these input costs. It is worth noting that a feature of the analysis presented in this paper is that we do not make any use of distance functions or the concept of technical efficiency by itself (although technical inefficiency can be a part of the cost efficiency concept defined by (4) below). The only functions that we use are the industry best practice cost function and linear functions that use the observed prices and quantities that pertain to a production unit for two periods.

[6] Notation: $w >> 0_N$ means that each element of the N dimensional vector w is positive, $w \geq 0_N$ means that each element of w is nonnegative and $w > 0_N$ means $w \geq 0_N$ but $w \neq 0_N$. The inner product of the N dimensional vectors $w \equiv [w_1,...,w_N]$ and $x \equiv [x_1,...,x_N]$ is denoted as $w \cdot x \equiv \sum_{n=1}^{N} w_n x_n$.

output and input vectors, (y,x), that could be produced in period t if the production unit had access to this technology. In the Appendix, we list some minimal regularity properties that we assume that the sets S^t possess for t = 0,1. We say that there has been *technical progress* between periods 0 and 1 if the production possibilities set S^1 is bigger than the corresponding period 0 set S^0 so that S^0 is a strict subset of S^1. Our final assumption for this section is that we can somehow solve the particular production unit's cost minimization problem for each period using the best practice technology. Thus for $y \geq 0_M$ and $w >> 0_N$, define the *period t best practice cost function*, $C^t(y,w)$, as follows:

$$(2)\ C^t(y,w) \equiv \min_x \{w \cdot x : (y,x) \in S^t\}\ ; \qquad t = 0,1.$$

Under our minimal regularity conditions on the production possibilities set S^t, it can be shown that $C^t(y,w)$ will be a nonnegative function, defined for all $y \geq 0_M$ and $w >> 0_N$, nondecreasing in the components of y for fixed w and concave, continuous, linearly homogeneous and nondecreasing in the components of w for fixed y. If we place stronger regularity conditions on the best practice technology, then $C^t(y,w)$ will satisfy stronger regularity conditions.[7]

In order to implement the decompositions that will be developed in this and subsequent sections, it is necessary that the analyst have estimates of the best practice cost functions, $C^t(y,w)$, for periods 0 and 1. This is possible in the context of a panel DEA study or in the context of estimating an econometric cost function[8] or a stochastic frontier production function using panel data for production units engaged in similar activities.

It will not necessarily be the case that the production unit being studied achieves the best practice level of costs; i.e., the following inequalities will be satisfied:

$$(3)\ c^t = w^t \cdot x^t \geq C^t(y^t,w^t)\ ; \qquad t = 0,1.$$

Thus the observed period t cost for the unit, c^t, will be equal to or greater than the best practice minimum cost, $C^t(y^t,w^t)$, where this minimum cost is computed using the period t best practice technology, the same vector of outputs y^t that the unit produced during period t and facing the same input prices w^t that the production unit faced during period t. Obviously, the difference between these two costs or their ratio can serve as a measure of the cost efficiency of the unit during period t. We will find it convenient to work with the ratio concept and thus we define the *cost efficiency* of the production unit during period t, e^t, as follows:[9]

$$(4)\ e^t \equiv C^t(y^t,w^t)/w^t \cdot x^t \leq 1\ ; \qquad t = 0,1$$

where the inequalities in (4) follow from (3). Thus if the establishment or firm is *cost efficient* in period t, e^t will equal its upper bound of 1. Note that the above definition of

[7] Thus if we assume that S^t is a closed convex cone, then $C^t(y,w)$ will be a linearly homogeneous, convex and nondecreasing function of y for fixed w.
[8] For an example of such an econometric study, see Lawrence and Diewert (2006).
[9] Balk (1998; 28) makes extensive use of this definition.

cost efficiency is equivalent to Farrell's (1957; 255) measure of *overall efficiency* in the DEA context, which combines his measures of technical and (cost) allocative efficiency. DEA or *Data Envelopment Analysis* is the term used by Charnes and Cooper (1985) and their co-workers to denote an area of analysis which is called the nonparameteric approach to production theory[10] or the measurement of the efficiency of production[11] by economists.

Given the above definition of cost efficiency in period t, we can define an index of the *change in the production unit's cost efficiency* over the two periods as follows:

(5) $\varepsilon(e^0,e^1) \equiv e^1/e^0 = [C^1(y^1,w^1)/w^1\cdot x^1]/[C^0(y^0,w^0)/w^0\cdot x^0]$ using (4).

Thus if $\varepsilon(e^0,e^1) > 1$, then the cost efficiency of the production unit has *improved* going from period 0 to 1 whereas it has *fallen* if $\varepsilon(e^0,e^1) < 1$.

Notice that the best practice (joint) cost function for period t, $C^t(y,w)$, depends on three sets of variables:

- The time period t and this index t serves to indicate that the period t best practice technology set S^t is being used by the production unit;
- The vector of outputs y which is to be produced by a best practice production unit and
- The vector of input prices w that the production unit faces.

At this point, we will follow the methodology that is used in the theoretical index number literature that originated with Konüs (1939) and Allen (1949) and we will use the cost function to define *three families of indexes* that vary only one of the three sets of variables, t, y and w, between the two periods under consideration and hold constant the other two sets of variables.[12]

Our first family of factors that explain cost changes is a *family of cost based measures of output change*, $\alpha(y^0,y^1,w,t)$:

(6) $\alpha(y^0,y^1,w,t) \equiv C^t(y^1,w)/C^t(y^0,w)$

Thus the effects on best practice cost due to a change in outputs is equal to the (hypothetical) total cost $C^t(y^1,w)$ of producing the vector of observed period 1 outputs for the production unit, y^1, divided by the total cost $C^t(y^0,w)$ of producing the vector of

[10] See Hanoch and Rothschild (1972), Diewert and Parkan (1983), Varian (1984) and Diewert and Mendoza (2007).
[11] See Farrell (1957), Afriat (1972), Färe and Lovell (1978), Färe, Grosskopf and Lovell (1985), Coelli, Rao and Battese (1997) and Balk (1998) (2003).
[12] The theory which follows is adapted from Diewert (2011). This approach to the output quantity and input price indexes is a reasonably straightforward adaptation of the earlier work on theoretical price and quantity indexes by Konüs (1939), Allen (1949), Fisher and Shell (1972), Samuelson and Swamy (1974), Archibald (1977), Diewert (1980; 461) (1983; 1054-1083) and Balk (1998).

observed period 0 outputs, y^0, where in both cases, we use the best practice technology of period t and assume that the production unit faces the vector of reference input prices, $w >> 0_N$. 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 measure of the effects on best practice cost of a change in output quantities.

Members of the family of output quantity change measures defined by (6) will not in general satisfy all of the properties of an output index that were considered desirable by O'Donnell (2008; 3).[13] Some of these desirable properties for an output index are:

- *Identity*; i.e., $\alpha(y^0,y^1,w,t) = 1$ if $y^0 = y^1$;
- *Homogeneity*; i.e., $\alpha(y^0,\lambda y^1,w,t) = \lambda\alpha(y^0,y^1,w,t)$ and $\alpha(\lambda y^0,y^1,w,t) = \lambda^{-1}\alpha(y^0,y^1,w,t)$ for all $y^0 >> 0_M$, $y^1 >> 0_M$ and $\lambda > 0$;
- *(Weak) Monotonicity*; i.e., if $y^2 \geq y^1 \geq 0_M$ and $y^0 >> 0_M$, then $\alpha(y^0,y^2,w,t) \geq \alpha(y^0,y^1,w,t)$ and
- *Continuity*; i.e., $\alpha(y^0,y^1,w,t)$ is a continuous function of y^0,y^1 for $y^0 >> 0_M$ and $y^1 >> 0_M$.

In general, $\alpha(y^0,y^1,w,t)$ will satisfy only the identity and weak monotonicity properties with our weak regularity conditions on the technology sets S^t.[14]

Following the example of Konüs (1939), it is natural to single out two special cases of the family of output quantity indexes defined by (6): one choice where we use the period 0 technology and set the reference prices equal to the period 0 input prices w^0 and another choice where we use the period 1 technology and set the reference prices equal to the period 1 input prices w^1. It is these choices that are the most relevant for the production unit whose costs for the two periods are being compared. Thus define these special cases as α_0 and α_1:

(7) $\alpha_0 \equiv \alpha(y^0,y^1,w^0,0) = C^0(y^1,w^0)/C^0(y^0,w^0)$;
(8) $\alpha_1 \equiv \alpha(y^0,y^1,w^1,1) = C^1(y^1,w^1)/C^1(y^0,w^1)$.

Thus the cost based measure of output quantity change α_0 is a *Laspeyres type measure* that uses the period 0 technology and period 0 input prices w^0 as reference input prices and the cost based measure of quantity change α_1 is a *Paasche type measure* that uses the period 1 technology and period 1 input prices w^1 as reference input prices. Since both measures of output change, α_0 and α_1, are equally representative, a single estimate of cost change due to output quantity changes between the two periods should be set equal to a symmetric average of these two estimates. We will choose the geometric mean as our

[13] Note that the measures defined by (6) are not used in our preferred decomposition, (34) below.
[14] However, if we make the stronger assumptions that the sets S^t are closed convex cones, then $\alpha(y^0,y^1,w,t)$ will satisfy all of the above desirable properties.

preferred symmetric average[15] and thus our preferred measure of cost based output quantity growth is the following *Fisher type theoretical measure*, α_F:

(9) $\alpha_F \equiv [\alpha_0\alpha_1]^{1/2}$.

We now turn our attention to measures of the effects on best practice cost of input price change. We use the period t best practice cost function C^t in order to define a *family of input price indexes*, $\beta(w^0,w^1,y,t)$, as follows:[16]

(10) $\beta(w^0,w^1,y,t) \equiv C^t(y,w^1)/C^t(y,w^0)$.

Thus the input price index $\beta(w^0,w^1,y,t)$ defined by (10) is equal to the (hypothetical) total cost $C^t(y,w^1)$ of producing the reference vector of outputs, $y >> 0_M$ when a production unit faces 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 production unit faces the period 0 observed vector of input prices w^0, where in both cases, the production unit has access to the best practice 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 (possibly different) input price index. Since the best practice cost function, $C^t(y,w)$, is linearly homogeneous, nondecreasing and concave in the components of w, it can be verified that each input price index $\beta(w^0,w^1,y,t)$ of the type defined by (10) will satisfy price counterparts to the identity, homogeneity, weak monotonicity and continuity properties listed above.[17]

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 (10): one choice where we use the period 0 technology and set the reference quantities equal to the period 0 quantities y^0 (which gives rise to a *Laspeyres type input price index*) and another choice where we use the period 1 technology and set the reference quantities equal to the period 1 quantities y^1 (which gives rise to a *Paasche type input price index*). Thus define these special cases as β_0 and β_1:

(11) $\beta_0 \equiv \beta(w^0,w^1,y^0,0) = C^0(y^0,w^1)/C^0(y^0,w^0)$;
(12) $\beta_1 \equiv \beta(w^0,w^1,y^1,1) = C^1(y^1,w^1)/C^1(y^1,w^0)$.

Since both 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

[15] Diewert (1997) explained why the geometric mean is a good choice for the symmetric average. Basically, the resulting indexes using geometric averages satisfy appropriate time reversal tests whereas other simple averages do not have this desirable property.
[16] If the number of outputs is equal to one, then the family of indexes defined by (10) reduces to the Konüs (1939) true cost of living index family where C^t is the consumer's expenditure function and output is interpreted as a utility level.
[17] The continuity of $\beta(w^0,w^1,y,t)$ in w^0 and w^1 (for $w^0 >> 0_N$ and $w^1 >> 0_N$) follows from the fact that a concave function is continuous in the interior of its domain of definition.

preferred overall measure of input price growth is the following *Fisher type theoretical index*, β_F:

(13) $\beta_F \equiv [\beta_0\beta_1]^{1/2}$.

We now define our last family of cost function based indexes for this section. We again use the total cost function in order to define a *family of indexes of technical progress*, $\tau(y,w)$, for a reference vector of outputs $y >> 0_M$ and a reference vector of input prices $w >> 0_N$ as follows:[18]

(14) $\tau(y,w) \equiv C^0(y,w)/C^1(y,w)$.

Technical progress measures are usually defined in terms of upward shifts in production functions or outward shifts of production possibilities sets due to the discovery of new techniques or managerial innovations over time. However, in the regulatory literature, it is quite common to specify technical progress in terms of downward shifts in the cost function over time. Thus in definition (14), we pick reference vectors y and w and use the best practice technology of period 0 to calculate the minimum cost of producing the output vector y at the input prices w using the period 0 and 1 best practice technologies at those times. This gives rise to the total costs, $C^0(y,w)$ and $C^1(y,w)$, respectively. If there is positive technical progress going from period 0 to 1, then $C^1(y,w)$ will be less than $C^0(y,w)$ and hence $\tau(y,w) = C^0(y,w)/C^1(y,w)$ will be greater than one and this measure of technical progress is the reciprocal of the degree of proportional cost reduction that results from the expansion of the underlying best practice technology sets due to the passage of time. For each choice of a reference vector of output quantities y and reference vector of input prices w, we obtain a (possibly different) measure of exogenous cost reduction and hence of technical progress.[19]

Instead of singling out the reference vectors y and w that appear in the definition of $\tau(y,w)$ 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 in more detail but basically, we make these choices in order to have cost decompositions into explanatory factors that are exact.

(15) $\tau_{01} \equiv \tau(y^0,w^1) = C^0(y^0,w^1)/C^1(y^0,w^1)$;
(16) $\tau_{10} \equiv \tau(y^1,w^0) = C^0(y^1,w^0)/C^1(y^1,w^0)$.

[18] This cost based measure of technical progress can be traced back to Salter (1960). For an early application of this measure, see Førsund and Hjalmarsson (1983). Balk (1998; 58) defined the family of indexes (14) in complete generality.
[19] In order to obtain a measure of technical progress that is invariant to the choice of y and w, it is necessary to make further restrictions on the technology sets, S^0 and S^1. However, we do not want to limit the generality of our model by restricting the best practice technology sets. Instead, we will pick reference y's and w's that are relevant or representative for the production unit. This means that we should pick y to be y^0 or y^1 and w to be w^0 or w^1.

Since both of the above measures of technical progress, τ_{01} and τ_{10}, are equally representative, a single estimate of technical progress should be set equal to a symmetric average of these two estimates. We again choose the geometric mean as our preferred symmetric average[20] and thus our preferred summary measure of technical progress going from period 0 to 1 is the following *Fisher type index of technical progress*, τ_F:

$$(17)\ \tau_F \equiv [\tau_{01}\,\tau_{10}]^{1/2}.$$

We want to explain the growth in total costs going from period 0 to 1 for the production unit under consideration, $c^1/c^0 = w^1\cdot x^1/w^0\cdot x^0$, as the product of four growth factors:

- Growth in outputs; i.e., a factor of the form $\alpha(y^0,y^1,w,t)$ defined above by (6);
- Growth in input prices; i.e., a factor of the form $\beta(w^0,w^1,y,t)$ defined by (10);
- Exogenous reduction in costs due to technical progress; i.e., a factor of the form $\tau(y,w)$ defined by (14) and
- Changes in the production unit's cost efficiency over the two periods; i.e., a factor of the form $\varepsilon(e^0,e^1)$ defined by (5) above.

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:[21]

$$\begin{aligned}(18)\ w^1\cdot x^1/w^0\cdot x^0 &= \{[w^1\cdot x^1/w^0\cdot x^0]/[C^1(y^1,w^1)/C^0(y^0,w^0)]\}\{C^1(y^1,w^1)/C^0(y^0,w^0)\} \\ &= [C^1(y^1,w^1)/C^0(y^0,w^0)]/\varepsilon(e^0,e^1) && \text{using definition (5)} \\ &= [C^1(y^1,w^1)/C^1(y^0,w^1)][C^0(y^0,w^1)/C^0(y^0,w^0)][C^1(y^0,w^1)/C^0(y^0,w^1)]/\varepsilon(e^0,e^1) \\ &= \alpha_1\beta_0/[\tau_{01}\varepsilon(e^0,e^1)] && \text{using definitions (8), (11) and (15);}\end{aligned}$$

$$\begin{aligned}(19)\ w^1\cdot x^1/w^0\cdot x^0 &= \{[w^1\cdot x^1/w^0\cdot x^0]/[C^1(y^1,w^1)/C^0(y^0,w^0)]\}\{C^1(y^1,w^1)/C^0(y^0,w^0)\} \\ &= [C^1(y^1,w^1)/C^0(y^0,w^0)]/\varepsilon(e^0,e^1) && \text{using definition (5)} \\ &= [C^0(y^1,w^0)/C^0(y^0,w^0)][C^1(y^1,w^1)/C^1(y^1,w^0)][C^1(y^1,w^0)/C^0(y^1,w^0)]/\varepsilon(e^0,e^1) \\ &= \alpha_0\beta_1/[\tau_{10}\,\varepsilon(e^0,e^1)] && \text{using definitions (7), (12) and (16).}\end{aligned}$$

Thus we have two equally valid exact decompositions of observed cost growth for the production unit given by (18) and (19) and if we take the geometric means of the two sides of (18) and (19), we obtain the following (preferred) *third exact decomposition of cost growth* using definitions (9), (13) and (17):

$$(20)\ w^1\cdot x^1/w^0\cdot x^0 = \alpha_F\beta_F/[\tau_F\,\varepsilon(e^0,e^1)].$$

Thus for each of the three decompositions (18)-(20), we have observed cost growth explained by the product of an output growth measure times a measure of input price

[20] This will ensure that the resulting measure of technical progress satisfies the time reversal property; i.e., if we reverse the role of time and recalculate the measure of technical progress, we obtain the reciprocal of the original measure when we take the geometric average.

[21] Diewert (2011) obtained decompositions of cost growth similar to (18) and (19) under the assumption that the production unit was cost efficient in each period.

growth, divided by the product of a measure of technical progress growth times a measure of growth in the cost efficiency of the production unit over the two time periods under consideration. Our preferred decomposition is (20) because it is a geometric average of the decompositions (18) and (19) and hence will generally be more representative or reliable than either of these measures by themselves.

In order to calculate the explanatory factors on the right hand sides of the decompositions of cost growth given by (18)-(20), it is necessary to have at hand estimates of the best practice cost functions for each period, $C^0(y,w)$ and $C^1(y,w)$, for y and w in a neighbourhood of the observed data, (y^0,w^0) and (y^1,w^1). In the following section, we assume that additional information on the production unit is available: namely information on the unit's selling prices for outputs in the two periods. With this additional information, we can look at decompositions of the growth in profits into explanatory factors for the production unit.

3. Profitability Growth Decompositions

We now assume that we can observe strictly positive output prices $p^t >> 0_M$ for the production unit in period t for t = 0,1. Thus in addition to having estimates of the production unit's period t *costs*, $c^t \equiv w^t \cdot x^t$, we now have estimates for the corresponding period t *revenues*, $r^t \equiv p^t \cdot y^t$ for t = 0,1. Hence, estimates of the production unit's period t *profitability* π^t [22] can be formed:

$$(21)\ \pi^t \equiv p^t \cdot y^t / w^t \cdot x^t\ ; \qquad t = 0,1.$$

In this section, we will attempt to find explanatory decompositions for the production unit's *profitability growth rate* (or *margin growth rate*), π_G, defined as follows:

$$(22)\ \pi_G \equiv \pi^1/\pi^0 = [p^1 \cdot y^1/w^1 \cdot x^1]/[p^0 \cdot y^0/w^0 \cdot x^0] = [p^1 \cdot y^1/p^0 \cdot y^0]/[w^1 \cdot x^1/w^0 \cdot x^0].$$

Thus the profitability growth rate is also equal to the revenue growth rate, r^1/r^0, divided by the cost growth rate, c^1/c^0.[23]

As in section 2, we will use the best practice cost functions, $C^t(y,w)$, in order to derive many of our explanatory factors for the rate of growth of profitability. All of the concepts that we need to provide our decomposition for π_G have been introduced in section 2 with two exceptions: we need to define some output price indexes and some indexes of returns to scale.

---

[22] The term "profitability" to describe this concept was introduced by Balk (2003; 9-10) but he considered the concept earlier; see Balk (1998; 66) for historical references. Diewert and Nakamura (2003; 129) described the same concept by the term "margin"; i.e., they set $\pi^t = 1+m^t$ where m^t is the period t *margin* for the production unit and they described π^1/π^0 as the "margin growth rate".

[23] There is a slight abuse of language here; what we are calling growth rates are actually one plus growth rates. Of course, growth rates can be positive or negative.

The output price indexes that we will need are quite conventional. The Laspeyres, Paasche and Fisher ideal *output price indexes*, P_L, P_P and P_F are defined as follows:

(23) $P_L(p^0,p^1,y^0,y^1) \equiv p^1\cdot y^0/p^1\cdot y^0$;
(24) $P_P(p^0,p^1,y^0,y^1) \equiv p^1\cdot y^1/p^1\cdot y^1$;
(25) $P_F(p^0,p^1,y^0,y^1) \equiv [P_L P_P]^{1/2}$.

A family of *returns to scale measures* ρ for the best practice technology of period t using the output vectors y^0 and y^1 that the production unit produced during periods 0 and 1, using the reference price vector $p >> 0_M$ to aggregate outputs and using the period t best practice cost function $C^t(y,w)$ (using the reference input price vector w) to aggregate the inputs used to produce y is defined as follows:[24]

(26) $\rho(y^0,y^1,p,w,t) \equiv [p\cdot y^1/p\cdot y^0]/[C^t(y^1,w)/C^t(y^0,w)]$.

Thus the returns to scale measure is equal to a measure of output growth using the reference output price vector p to aggregate the outputs, $p\cdot y^1/p\cdot y^0$, divided by the corresponding aggregate input cost growth measure, $C^t(y^1,w)/C^t(y^0,w)$, which is equal to the input cost ratio using the best practice technology for period t and the reference input price vector w to aggregate the inputs.[25] The period t best practice technology exhibits *increasing returns to scale* as we move from the output vector y^0 to the output vector y^1 if $\rho(y^0,y^1,p,w,t) > 1$ so that the output aggregate grows more rapidly than the input aggregate, *decreasing returns to scale* if $\rho(y^0,y^1,p,w,t) < 1$ and *constant returns to scale* if $\rho(y^0,y^1,p,w,t) = 1$.[26] As usual, there are two special cases of (26) that are of interest to us:

---

[24] This measure of returns to scale is not quite conventional (except when the number of outputs equals one) since it combines a conventional measure of returns to scale with mix effects. A conventional (global) measure of returns to scale is similar to definition (26) except y^1 is restricted to equal λy^0 for some scalar $\lambda > 0$. In order to define a conventional measure of returns to scale, first define λ^1 by the following equation: $p\cdot\lambda^1 y^0 = p\cdot y^1$ or $\lambda^1 = p\cdot y^1/ p\cdot y^0$. Thus $\lambda^1 y^0$ is a scaled up version of y^0 that gives the same output revenue as y^1 using the reference prices p. Using this λ^1, the conventional definition of returns to scale is defined as $\rho^*(y^0,\lambda^1 y^0,p,w,t) \equiv [p\cdot\lambda^1 y^0/p\cdot y^0]/[C^t(\lambda^1 y^0,w)/C^t(y^0,w)] = [p\cdot y^1/p\cdot y^0]/[C^t(\lambda^1 y^0,w)/C^t(y^0,w)]$. The effects on best practice cost of shifting from producing the output vector y^1 to producing $\lambda^1 y^0$ (note that both output vectors generate the same revenue at the reference prices p) is the *output mix adjustment factor* $\mu(\lambda^1 y^0,y^1,p,w,t) \equiv C^t(\lambda^1 y^0,w)/C^t(y^1,w)$ where $\lambda^1 \equiv p\cdot y^1/p\cdot y^0$. Using these definitions, $\rho(y^0,y^1,p,w,t) = \rho^*(y^0,\lambda^1 y^0,p,w,t)\ \mu(\lambda^1 y^0,y^1,p,w,t)$. For additional material on cost function based definitions of returns to scale and mix effects, see Balk (2001).

[25] Our measure of returns to scale is a "global" one as opposed to a "local" one. A local definition of returns to scale for the best practice cost function will be given in the following section. This local definition for returns to scale requires that the cost function $C^t(y,w)$ be differentiable with respect to the components of y. We have tried to avoid making differentiability assumptions on C^t because we want our analysis to be applicable to the type of nondifferentiable cost functions that arise in the DEA literature.

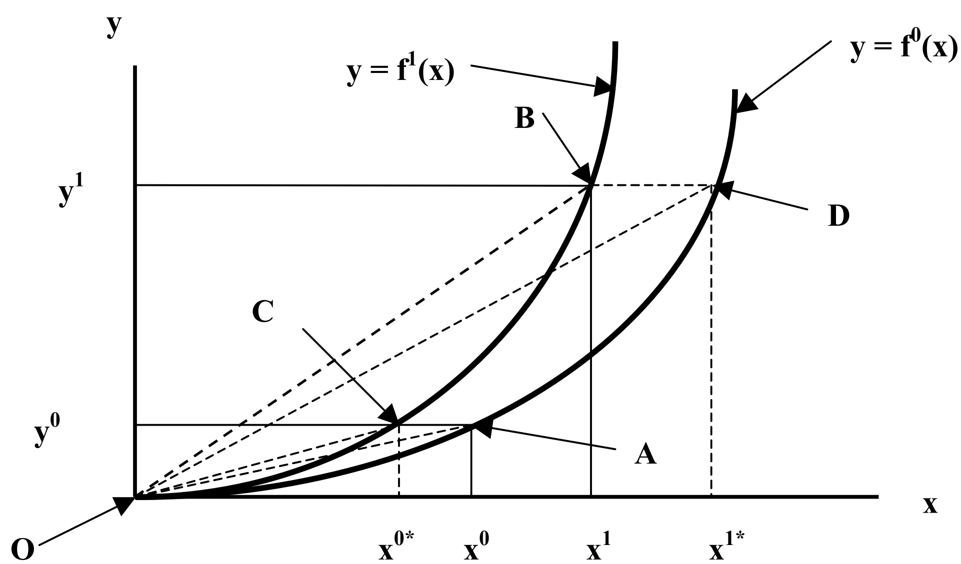
[26] It is possible to give an alternative interpretation for ρ as follows. Define the hypothetical markup profitability m(y,p,w,t) for a production unit as a function of the output vector y for the reference output price vector p and input price vector w using the best practice technology for period t as $m(y,p,w,t) \equiv p\cdot y/C^t(y,w)$. Then the ratio of profitability producing output y^1 to profitability producing output y^0 is $m(y^1,p,w,t)/m(y^0,p,w,t) = \rho(y^0,y^1,p,w,t)$. Thus ρ can be interpreted as a *measure of profitability change due to changes in the vector of outputs produced* using the best practice technology of period t at the reference output and input price vectors, p and w.

one case where we set the reference variables equal to their base period values (which gives rise to the returns to scale measure ρ_0) and another case where we set the reference variables equal to their period 1 values (which gives rise to the returns to scale measure ρ_1):[27]

(27) $\rho_0 \equiv \rho(y^0,y^1,p^0,w^0,0) = [p^0 \cdot y^1/C^0(y^1,w^0)]/[p^0 \cdot y^0/C^0(y^0,w^0)]$;
(28) $\rho_1 \equiv \rho(y^0,y^1,p^1,w^1,1) = [p^1 \cdot y^1/C^1(y^1,w^1)]/[p^1 \cdot y^0/C^1(y^0,w^1)]$.

It is possible to give a graphical interpretation of the returns to scale measures defined by (27) and (28) in the case of one output and one input. In this case, the cost function for period t is equal to the inverse of the period t production function times the input price so if $y = f^t(x)$ represents the production function, then $x = g^t(y)$ is the inverse (factor requirements) function and the period t cost function is equal to $C^t(y,w) = wg^t(y)$.[28]

Figure 1: Returns to Scale Measures in the One Output, One Input Case



For simplicity, we assume that production is efficient in both periods and the points A and B in Figure 1 represent the observed data points for the production unit in the two periods under consideration, (y^0,x^0) and (y^1,x^1). Using the period 0 production function, it would take x^{1*} units of input to produce the output y^1 of period 1; i.e., we have $y^1 = f^0(x^{1*})$ or $x^{1*} = g^0(y^1)$. Thus the period 0 returns to scale measure, ρ_0, is equal to:

(29) $\rho_0 = [p^0y^1/w^0x^{1*}]/[p^0y^0/w^0x^0] = [y^1/x^{1*}]/[y^0/x^0]$

---

[27] ρ_0 can be replaced by $\rho^*(y^0,\lambda^1y^0,p^0,w^0,0)\ \mu(\lambda^1y^0,y^1,p^0,w^0,0) \equiv \rho_0^*\ \mu_0$ where $\lambda^1 = p^0 \cdot y^1/\ p^0 \cdot y^0$ and ρ_1 can be replaced by $\rho^*(y^0,\lambda^1y^0,p^1,w^1,1)\ \mu(\lambda^1y^0,y^1,p^1,w^1,1) \equiv \rho_1^*\ \mu_1$ where $\lambda^1 = p^1 \cdot y^1/\ p^1 \cdot y^0$.

[28] In order to simplify the notation in the diagram, we write y_1, x_1 and w_1 as y, x and w, etc.

which is the slope of the line segment OD divided by the slope of the line segment OA. Thus in this case, we have $\rho_0 > 1$ so that the period 0 technology over the relevant range of outputs produced in the two periods is subject to increasing returns to scale. In a similar fashion, it can be seen that using the period 1 production function, it would take x^{0*} units of input to produce the output y^0 of period 0; i.e., we have $y^0 = f^1(x^{0*})$ or $x^{0*} = g^1(y^0)$. Thus the period 1 returns to scale measure, ρ_1, is equal to:

$$(30)\ \rho_1 = [p^1y^1/w^1x^1]/[p^1y^0/w^1x^{0*}] = [y^1/x^1]/[y^0/x^{0*}]$$

which is the slope of the line segment OB divided by the slope of the line segment OC. Thus in this case, we have $\rho_1 > 1$ so that the period 1 technology over the relevant range of outputs produced in the two periods is also subject to increasing returns to scale.

As usual, we will find it convenient to define an average of the two returns to scale measures defined by (27) and (28):

$$(31)\ \rho_F \equiv [\rho_0\,\rho_1]^{1/2}.$$

We will now provide two equally plausible exact decompositions of profitability growth π_G and then, as usual, we will take the geometric mean of these two decompositions as our preferred decomposition. Using definition (22), we have:

$$\begin{aligned}(32)\ \pi_G &= [p^1\cdot y^1/p^0\cdot y^0]/[w^1\cdot x^1/w^0\cdot x^0]\\
&= [p^1\cdot y^0/p^0\cdot y^0][p^1\cdot y^1/p^1\cdot y^0][C^0(y^0,w^0)/C^1(y^1,w^1)][C^1(y^1,w^1)/w^1\cdot x^1]/[C^0(y^0,w^0)/w^0\cdot x^0]\\
&= [p^1\cdot y^0/p^0\cdot y^0][p^1\cdot y^1/p^1\cdot y^0][C^0(y^0,w^0)/C^1(y^1,w^1)]\varepsilon(e^0,e^1) \qquad \text{using definition (5)}\\
&= \varepsilon(e^0,e^1)[p^1\cdot y^0/p^0\cdot y^0]\{[p^1\cdot y^1/C^1(y^1,w^1)]/[p^1\cdot y^0/C^1(y^0,w^1)]\}[C^0(y^0,w^0)/C^1(y^0,w^1)]\\
&= \varepsilon(e^0,e^1)P_L(p^0,p^1,y^0,y^1)\rho_1[C^0(y^0,w^0)/C^1(y^0,w^1)] \qquad \text{using definitions (23) and (28)}\\
&= \varepsilon(e^0,e^1)P_L(p^0,p^1,y^0,y^1)\rho_1[C^0(y^0,w^1)/C^1(y^0,w^1)]/[C^0(y^0,w^1)/C^0(y^0,w^0)]\\
&= \varepsilon(e^0,e^1)\,P_L(p^0,p^1,y^0,y^1)\,\rho_1\,\tau_{01}/\beta_0 \qquad \text{using definitions (11) and (15).}\end{aligned}$$

Thus profitability growth, π_G, is exactly equal to the growth in cost efficiency of the production unit $\varepsilon(e^0,e^1)$ times the Laspeyres output price index $P_L(p^0,p^1,y^0,y^1)$ times the period 1 measure of returns to scale ρ_1 times the measure of best practice technical progress τ_{01} defined by (15) divided by the Laspeyres type input price index β_0 using the period 0 best practice technology defined above by (11).

In a similar fashion, we can derive the following decomposition for the rate of margin growth:

$$\begin{aligned}(33)\ \pi_G &= [p^1\cdot y^1/p^0\cdot y^0]/[w^1\cdot x^1/w^0\cdot x^0]\\
&= [p^1\cdot y^1/p^0\cdot y^1][p^0\cdot y^1/p^0\cdot y^0][C^0(y^0,w^0)/C^1(y^1,w^1)][C^1(y^1,w^1)/w^1\cdot x^1]/[C^0(y^0,w^0)/w^0\cdot x^0]\\
&= [p^1\cdot y^1/p^0\cdot y^1][p^0\cdot y^1/p^0\cdot y^0][C^0(y^0,w^0)/C^1(y^1,w^1)]\varepsilon(e^0,e^1) \qquad \text{using definition (5)}\\
&= \varepsilon(e^0,e^1)[p^1\cdot y^1/p^0\cdot y^1]\{[p^0\cdot y^1/C^0(y^1,w^0)]/[p^0\cdot y^0/C^0(y^0,w^0)]\}[C^0(y^1,w^0)/C^1(y^1,w^1)]\\
&= \varepsilon(e^0,e^1)P_P(p^0,p^1,y^0,y^1)\rho_0[C^0(y^1,w^0)/C^1(y^1,w^1)] \qquad \text{using definitions (24) and (27)}\\
&= \varepsilon(e^0,e^1)P_P(p^0,p^1,y^0,y^1)\rho_0[C^0(y^1,w^0)/C^1(y^1,w^0)]/[C^1(y^1,w^1)/C^1(y^1,w^0)]\end{aligned}$$

$= \varepsilon(e^0,e^1)\, P_P(p^0,p^1,y^0,y^1)\, \rho_0\, \tau_{10}/\beta_1$ using definitions (12) and (16).

Thus profitability growth, π_G, is exactly equal to the growth in cost efficiency of the production unit $\varepsilon(e^0,e^1)$ times the Paasche output price index $P_P(p^0,p^1,y^0,y^1)$ times the period 0 measure of returns to scale ρ_0 times the measure of best practice technical progress τ_{10} defined by (16) divided by the Laspeyres type input price index β_0 using the period 0 best practice technology defined above by (12).

By taking the geometric mean of the decompositions of margin growth defined by (32) and (33), we obtain our *preferred decomposition for profitability growth*:

(34) $\pi_G = \varepsilon(e^0,e^1)\, P_F(p^0,p^1,y^0,y^1)\, \rho_F\, \tau_F/\beta_F$

where $\varepsilon(e^0,e^1)$ is the growth in the production unit's cost efficiency defined by (5), $P_F(p^0,p^1,y^0,y^1)$ is the Fisher output price index defined by (25), ρ_F is the Fisher type measure of returns to scale defined by (31), τ_F is the Fisher type measure of technical progress for the best practice technology defined by (17) and β_F is the Fisher type measure of input price growth for the best practice technology defined by (17). Thus there are *five separate explanatory factors* that help to explain profitability growth.[29]

The effects of output price change and input price change in (34) can be combined into one term. Thus define the *output-input price index*[30] for the production unit, P_G, as follows:

(35) $P_G \equiv P_F(p^0,p^1,y^0,y^1)/\beta_F$.

Substituting (35) into (34) leads to the following decomposition of profitability growth for the production unit:

(36) $\pi_G = \varepsilon(e^0,e^1)\, \rho_F\, \tau_F\, P_G$.

The decomposition of profitability growth given by (36) can be compared to the alternative profitability growth decompositions obtained by Balk (2003; 22) and O'Donnell (2008; 11) (2010; 531) (2012).

Balk (2003; 22) noted that a decomposition like (34) could be useful in a regulatory context.[31] The regulator can fix the margin growth rate π_G to a suitable level (equal to

[29] Recalling footnote 26, it can be verified that ρ_F can be replaced by $\rho_F^* \equiv \rho_F^*\mu_F$ where $\rho_F^* \equiv (\rho_0^*\rho_1^*)^{1/2}$ and $\mu_F \equiv (\mu_0\mu_1)^{1/2}$. With this replacement, profitability growth is decomposed into six explanatory factors.
[30] O'Donnell (2008; 11) (2010; 531) used the description "terms of trade index" to describe this index. Balk (2003; 22) noted that P_G has been described as a "terms of trade index", a "price recovery index" or a "price performance index". The problem with the "terms of trade" description is that this term also has a well established meaning in the literature as the ratio of an export price index to an import price index. Thus to prevent confusion, we have used the rather mundane but descriptive term "output-input price index" to describe this concept. Note that our concept differs from Balk's (1998; 164) "simultaneous input and output price index".

unity if economic profits of the regulated production unit in period 0 are equal to 0) and then rearrange equation (34) to give the following formula for the Fisher output price index:

(37) $P_F(p^0,p^1,y^0,y^1) = \pi_G\,\beta_F/[\varepsilon(e^0,e^1)\rho_F\,\tau_F]$.

Thus output prices could be allowed to grow, according to a price cap given by the Fisher ideal index $P_F(p^0,p^1,y^0,y^1)$, which would be set equal to $\pi_G\beta_F/\varepsilon(e^0,e^1)\rho_F\tau_F$. Thus the regulator would have to have estimates of expected input price growth for the best practice technology β_F, for the degree of expected cost efficiency improvement for the unit $\varepsilon(e^0,e^1)$, for the degree of returns to scale for the unit using the best practice technology ρ_F, and for expected technical progress τ_F in order to implement the price cap.[32]

Comparing our decomposition of profitability growth given by (34) to our earlier cost growth decomposition given by (20), it can be seen that the ordinary Fisher output price index $P_F(p^0,p^1,y^0,y^1)$ has replaced the cost function based index of output change α_F and a new explanatory factor has appeared in (34) which was not present in (20): namely the returns to scale index, ρ_F. One might think that returns to scale would also be a relevant explanatory factor in the cost growth decomposition (20) but unfortunately, in order to work a returns to scale measure into (20), we require an independent method for forming an output aggregate (other than by using the best practice cost function to aggregate outputs, which was the aggregation method used in section 2). In the present section, we have assumed the existence of output prices p^t which can be used in order to form independent output aggregates and hence, the returns to scale explanatory factor, ρ_F, makes its appearance in (34).

The margin growth decomposition given by (34) can be rearranged to give us equation (38) below:

(38) $\varepsilon(e^0,e^1)\rho_F\,\tau_F = \pi_G\beta_F/P_F(p^0,p^1,y^0,y^1) \equiv TFP_G$;

i.e., we have defined a *dual expression for TFP growth* for the production unit, TFP_G, as being equal to the product of *margin growth* for the production unit, π_G, times our Fisher type cost function based *index of input price change*, β_F, divided by the ordinary *Fisher output price index*, $P_F(p^0,p^1,y^0,y^1)$. The equality in (38) tells us that this dual measure of TFP growth is equal to the product of efficiency growth $\varepsilon(e^0,e^1)$ times returns to scale ρ_F times the measure of best practice technical progress τ_F. In order to obtain an equivalent primal measure of TFP growth, replace π_G in (34) by revenue growth, r^1/r^0, divided by cost growth, c^1/c^0, for the production unit and rearrange the resulting equation as follows:

---

[31] "A regulation agency might use this expression as a vehicle for placing a bound on the average output price change by restricting a firm's profitability ratio to a prescribed value. Then the allowed changes in the output prices will be determined by the rate of change of the input prices corrected by the rate of TFP change." Bert M. Balk (2003; 22).

[32] Diewert, Lawrence and Fallon (2009) derive related price cap formulae with the added complication of sunk cost capital inputs.

(39) $\varepsilon(e^0,e^1)\rho_F\,\tau_F = [(r^1/r^0)/P_F(p^0,p^1,y^0,y^1)]/[(c^1/c^0)/\beta_F] = TFP_G$.

Note that the revenue ratio divided by the Fisher output price index for the production unit, $(r^1/r^0)/P_F(p^0,p^1,y^0,y^1)$, is equal to the Fisher output quantity index $Q_F(p^0,p^1,y^0,y^1)$ for the production unit and the production unit's observed cost ratio divided by the cost function based Fisher type input price index, $(c^1/c^0)/\beta_F$, is an implicit input quantity index for the production unit. Thus the primal formula for TFP growth is equal to the ratio of this output quantity index to the corresponding input quantity index.

Finally, we note that the first equation in (39) can be rearranged to give the following decomposition of nominal revenue growth of the production unit going from period 0 to 1:

(40) $r^1/r^0 = p^1\cdot y^1/p^0\cdot y^0 = [(c^1/c^0)/\beta_F]P_F(p^0,p^1,y^0,y^1)\varepsilon(e^0,e^1)\rho_F\,\tau_F$.

Thus nominal value of output growth for the production unit is equal to an index of input quantity growth $(c^1/c^0)/\beta_F$ times an index of output price growth $P_F(p^0,p^1,y^0,y^1)$ times an index of cost efficiency growth $\varepsilon(e^0,e^1)$ times a measure of returns to scale ρ_F times a measure of technical progress τ_F. The decomposition is similar in spirit to a decomposition of nominal output growth obtained by Kohli (1990) who used a parametric translog approach to describe the technology of the production unit.[33]

4. Maximum TFP Allocations of Resources

O'Donnell (2008; 13) (2010; 533) noted that if we have a suitable period t output aggregate for the production unit (say Y^t) and a suitable aggregator for inputs (say X^t), then it makes sense to look for the *maximum possible output per unit input*, Y^t/X^t, for the production unit, using the period t technology. Since Y^t/X^t is generally defined as a *productivity level*, what we are looking for is a maximum TFP allocation of resources for the production unit for period t.

In this section, we will define the period t output aggregate to be $Y^t \equiv p^t\cdot y$ and the period t input aggregate to be $X^t \equiv C^t(y,w^t)$. Thus in order to find the maximum possible TFP that is possible for the production unit in period t, we want to solve the following maximization problem:[34]

(41) $\max_y \{p^t\cdot y/C^t(y,w^t) : y \geq 0_M\}$.

[33] Diewert and Morrison (1986) derived results similar to those of Kohli; see also Fox and Kohli (1998). These authors assumed a constant returns to scale technology and efficiency in each period. Diewert and Fox (2008) (2010) also derived decompositions analogous to (40) using translog approaches without assuming constant returns to scale but they also assumed cost efficiency in each period.

[34] Balk (1998; 66) considered a more general version of this problem. For an early analysis of this problem, see Samuelson (1967). O'Donnell (2012) solves this problem for the case where the best practice cost function is defined using DEA techniques.

Thus we look for an optimal period t output vector, say y^{t*}, which maximizes the production unit's revenue $p^t \cdot y^{t*}$ to cost $C^t(y^{t*},w^t)$ ratio or its period t margin or profitability. There is no guarantee that a solution y^{t*} to (41) will exist[35] but we will suppose that such a solution exists with $y^{t*} >> 0_M$. If $C^t(y,w^t)$ is differentiable with respect to y at the point y^{t*}, then the following first order necessary conditions must be satisfied:

$$(42)\ p^t/C^t(y^{t*},w^t) - p^t \cdot y^{t*}\ \nabla_y C^t(y^{t*},w^t)/[C^t(y^{t*},w^t)]^2 = 0_M \text{ or}$$
$$(43)\ p^t = p^t \cdot y^{t*}\ \nabla_y C^t(y^{t*},w^t)/C^t(y^{t*},w^t) = [p^t \cdot y^{t*}/C^t(y^{t*},w^t)]\ \mu^{t*}$$

where the vector of optimal marginal costs is defined as

$$(44)\ \mu^{t*} \equiv \nabla_y C^t(y^{t*},w^t).$$

Thus (43) tells us that if an interior solution to (41) exists, then optimal marginal costs μ^{t*} should be proportional to output prices p^t.

Take the inner product of both sides of (43) with the optimal output vector y^{t*} and simplify the resulting equation which becomes:

$$(45)\ 1 = C^t(y^{t*},w^t)/y^{t*} \cdot \nabla_y C^t(y^{t*},w^t).$$

But the right hand side of (45) is the Panzar and Willig's (1977; 490) dual measure of local returns to scale of the period t technology[36] and so equation (45) tells us that at the period t maximizing level of productivity (if it exists), the measure of local returns to scale is unity; i.e., we have local constant returns to scale at the TFP maximizing choice of outputs y^{t*}. In the case of only one output, (45) becomes the familiar marginal cost $\partial C^t(y^{t*},w^t)/\partial y_1$ equals average cost, $C^t(y_1^{t*},w^t)/y_1^{t*}$, that characterizes a maximal productivity allocation of resources.

5. Conclusion

In this paper, we have obtained a decomposition of profitability growth (34) into explanatory factors that is analogous to existing decompositions in the literature. What is new in our decomposition is that it depends only on observable price and quantity data pertaining to a production unit over two periods and industry best practice cost functions for the two periods under consideration. If output price data are not available, we obtain a similar decomposition (20) of a production unit's cost ratio into four explanatory factors.

---

[35] If there are increasing returns to scale everywhere, there will be no finite solution to (74). Also, if the technology is convex and exhibits strict decreasing returns to scale, the optimal y may be at the origin and so there may be no positive solution to (74).

[36] Ohta (1974) showed that the primal measure of local returns to scale (using the production function) is equal to this dual cost function measure of local returns to scale in the case of one output and Panzar and Willig (1977) generalized this dual measure to the case of many outputs.

Our main result is equation (34), which decomposes *profitability growth* π_G into the product of four multiplicative factors divided by a fifth factor. The four multiplicative factors in the numerator are:

- $\varepsilon(e^0,e^1)$, a measure of the growth in the production unit's *cost efficiency* defined by (5);
- $P_F(p^0,p^1,y^0,y^1)$, the *Fisher output price index* defined by (25);
- ρ_F, the *Fisher type measure of returns to scale* defined by (31) and
- τ_F, the *Fisher type measure of technical progress* for the best practice technology defined by (17).

The factor in the denominator of the decomposition is β_F, the Fisher type measure of *input price growth* for the best practice technology defined by (17).

If a complete knowledge of the best practice cost functions for the production unit under consideration is not available, then approximations to the above indexes could be obtained using first order approximations to the various indexes described above; see Diewert (2011).

Finally, we showed how our preferred decomposition of profitability growth or margin growth, (34), could be rearranged to give us decompositions of Total Factor Productivity growth into explanatory growth factors.

Appendix: Cost Functions and Regularity Conditions on the Technology

Let S denote a generic production possibilities set and define the (best practice) cost function C(y,w) that corresponds to this technology for $y \geq 0_M$ and $w >> 0_N$, as follows:

(A1) $C(y,w) \equiv \min_x \{w \cdot x : (y,x) \in S\}$.

Consider the following *five properties for S*:

P1. S is a nonempty closed subset of the nonnegative orthant in Euclidean M+N dimensional space.

P2. For every $y \geq 0_M$, there exists an $x \geq 0_N$ such that $(y,x) \in S$.

The interpretation of P2 is that every finite output vector y is producible by a finite input vector x.

P3. $(y,x^1) \in S$, $x^2 \geq x^1$ implies $(y,x^2) \in S$.

Thus if S satisfies P3, then there is free disposability of inputs.

P4. $y > 0_M$ implies that $(y,0_N) \notin S$.

The interpretation of P4 is that zero amounts of all inputs cannot produce a positive output.

P5. $(y^1,x)\in S$, $0_M \le y^2 \le y^1$ implies $(y^2,x)\in S$.

Thus if the input vector x can produce the output vector y^1 and y^2 is equal to or less than y^1, then x can also produce the smaller vector of outputs, y^1 (free disposability of outputs).

Under the above minimal regularity conditions on the production possibilities set S, it can be shown that C(y,w) will be a nonnegative function, defined for all $y \ge 0_M$ and $w >> 0_N$, positive for $y > 0_M$ and $w >> 0_N$, nondecreasing in the components of y for fixed w and concave, continuous, linearly homogeneous and nondecreasing in the components of $w >> 0_N$ for fixed y.[37]

If we want C(y,w) to be increasing in the components of y (instead of the above weaker nondecreasing property), then we will require a stronger regularity condition on the technology in order to ensure this property.

One such stronger property is the following one:

P6: If $x^1\in\phi(y^1,w)$ where $y^1 \ge 0_M$ and $w >> 0_N$ and $y^2 > y^1$, then $(y^2,x^1)\notin S$.

The set valued function $\phi(y^1,w)$ which appears in P6 is the set of input vectors x which solve the cost minimization problem defined by (A1); i.e., if $x\in\phi(y^1,w)$, then $(y^1,x)\in S$ and $w\cdot x = C(y^1,w)$. Thus P6 says that if the input vector x^1 is a solution to the cost minimization problem defined by $C(y^1,w)$ and we increase the output vector from y^1 to y^2 which is strictly greater than y^1 in at least one component, then x^1 is no longer a feasible solution to the cost minimization problem defined by $C(y^2,w)$. Thus as we increase our output targets, the set of feasible input vectors becomes strictly smaller in a certain sense. The property P6 will ensure that C(y,w) is increasing in the components of y. However, the regularity properties P1-P6 do not rule out discontinuous behavior of C(y,w) with respect to the components of y. In order to rule out discontinuities, we will need additional assumptions; i.e., the production surface will have to be continuous with no flat spots.

If we place even stronger regularity conditions on the best practice technology, then C(y,w) will satisfy stronger regularity conditions.[38]

[37] For proofs of these properties, see McFadden (1978) or Diewert (1993; 107-123).

[38] Thus if we assume that S^t is a closed convex cone, then C(y,w) will be a linearly homogeneous, convex and nondecreasing function of y for fixed w. For proofs of these properties and references to the literature, see Diewert (1993; 167-171). Note that the mathematical properties of joint cost functions can be established in the same way that the properties of profit functions were established: a minimization problem is solved instead of a maximization problem.

References

Afriat, S.N. (1972), "Efficiency Estimation of Production Function", *International Economic Review* 13, 568-598.

Allen, R.D.G. (1949), "The Economic Theory of Index Numbers", *Economica* 16, 197-203.

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

Balk, B.M. (1998), *Industrial Price, Quantity and Productivity Indices*, Boston: Kluwer Academic Publishers.

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

Balk, B.M. (2003), "The Residual: On Monitoring and Benchmarking Firms, Industries and Economies with respect to Productivity", *Journal of Productivity Analysis* 20, 5-47.

Charnes, A. and W.W. Cooper (1985), "Preface to Topics in Data Envelopment Analysis", *Annals of Operations Research* 2, 59-94.

Coelli, T., D.S. Prasada Rao and G. Battese (1997), *An Introduction to Efficiency and Productivity Analysis*, Boston: Kluwer Academic Publishers.

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. (1993), "Duality Approaches to Microeconomic Theory", pp. 105-189 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

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. (2011), "Measuring Productivity in the Public Sector: Some Conceptual Problems", *Journal of Productivity Analysis* 36; 177-191.

Diewert, W.E. and K.J. Fox (2008), "On the Estimation of Returns to Scale, Technical Progress and Monopolistic Markups", *Journal of Econometrics* 145, 174-193.

Diewert, W.E. and K.J. Fox (2010), "Malmquist and Törnqvist Productivity Indexes: Returns to Scale and Technical Progress with Imperfect Competition", *Journal of Economics* 101:1, 73-95.

Diewert, W.E., D. Lawrence and J. Fallon (2009), *The Theory of Network Regulation in the Presence of Sunk Costs,* Economic Insights, Report Prepared for Commerce Commission, Government of New Zealand, Wellington NZ, 2009. http://www.econ.ubc.ca/diewert/insights2.pdf

Diewert, W.E. and N.F. Mendoza (2007), "The Le Chatelier Principle in Data Envelopment Analysis", pp. 63-82 in *Aggregation, Efficiency, and Measurement*, Rolf Färe, Shawna Grosskopf and Daniel Primont (eds.), New York. http://www.econ.ubc.ca/diewert/chatelier.pdf

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.

Diewert, W.E. and A.O. Nakamura (2003), "Index Number Concepts, Measures and Decompositions of Productivity Growth", *Journal of Productivity Analysis* 19, 127-159.

Diewert, W.E. and C. Parkan (1983), "Linear Programming Tests of Regularity Conditions for Production Functions," pp. 131-158 in *Quantitative Studies on Production and Prices*, W. Eichhorn, R. Henn, K. Neumann and R.W. Shephard (eds.), Vienna: Physica Verlag.

Färe, R. and C.A.K. Lovell (1978), "Measuring the Technical Efficiency of Production", *Journal of Economic Theory* 19, 150-162.

Färe, R., S. Grosskopf and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff.

Farrell, M.J. (1957), "The Measurement of Production Efficiency", *Journal of the Royal Statistical Society*, Series A, 120, 253-278.

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.

Førsund, F. R. and Hjalmarsson, L. (1983), "Technological Progress and Structural Change in the Swedish Cement Industry 1955-1979", *Econometrica* 51, 1449-1467.

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.

Hanoch, G. and M. Rothschild (1972), "Testing the Assumptions of Production Theory: A Nonparametric Approach", *Journal of Political Economy* 80, 256-275.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

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

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.

McFadden, D. (1978), "Cost, Revenue and Profit Functions", pp. 3-109 in *Production Economics: A Dual Approach to Theory and Applications*, Volume 1, M. Fuss and D. McFadden (eds.), Amsterdam: North-Holland.

O'Donnell, C.J. (2008), "An Aggregate Quantity-Price Framework for Measuring and Decomposing Productivity and Profitability Change", Working Paper WP07/2008, Centre for Efficiency and Productivity Analysis, School of Economics, University of Queensland, St. Lucia, Queensland 4072, Australia, revised February 2009.

O'Donnell, C.J. (2010), "Measuring and Decomposing Agricultural Productivity and Profitability Change", *Australian Journal of Agricultural and Resource Economics* 54:4, 527-560.

O'Donnell, C.J. (2012), "Nonparametric Estimates of the Components of Productivity and Profitability Change in U.S. Agriculture", *American Journal of Agricultural Economics*, forthcoming.

Ohta, M. (1974), "A Note on the Duality Between Production and Cost Functions: Rate of Returns to Scale and Rate of Technical Progress", *Economic Studies Quarterly* 25, 63-65.

Panzar, J.C. and R.D. Willig (1977), "Economies of Scale in Multi-Output Production", *Quarterly Journal of Economics* 91, 481-493.

Samuelson, P.A. (1967), "The Monopolistic Competition Revolution", pp. 105-138 in *Monopolistic Competition Theory: Studies in Impact*, R.E. Kuenne (ed.), New York: John Wiley.

Schreyer, P. (2008), "Output and Outcome—Measuring the Production of Non-Market Services", OECD Statistics Directorate, April 15, 2008 draft.

Salter, W. E. G. (1960), *Productivity and Technical Change*, Cambridge U.K.: Cambridge University Press.

Varian, H.R. (1984), "The Nonparametric Approach to Production Analysis", *Econometrica* 52, 579-597.