Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP07/2008

Title
An aggregate quantity-price framework for measuring and
Decomposing productivity and profitability change

Authors
C.J. O'Donnell

Date: 13 November 2008
Revised February 2009

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

AN AGGREGATE QUANTITY-PRICE FRAMEWORK FOR MEASURING AND DECOMPOSING PRODUCTIVITY AND PROFITABILITY CHANGE[1]

by

C.J. O'Donnell

The University of Queensland
Centre for Efficiency and Productivity Analysis
Brisbane 4072, Australia
email: c.odonnell@economics.uq.edu.au

2 February 2009

Abstract: Total factor productivity (TFP) is often defined as the ratio of an aggregate output to an aggregate input. This definition naturally leads to TFP indexes that can be expressed as the ratio of an output quantity index to an input quantity index. In this paper, such index numbers are said to be *multiplicatively complete*. Complete indexes can be shown to satisfy important axioms from index number theory. This paper formally defines what is meant by completeness and demonstrates that i) the class of complete TFP index numbers includes Törnqvist, Konus and Moorsteen-Bjurek indexes, ii) the popular Malmquist TFP index of Caves, Christensen and Diewert (1982a, p.45) is incomplete, iii) any complete TFP index can be decomposed into measures of technical change, technical efficiency change, mix efficiency change, and scale efficiency change, and iv) profitability change can be broken into these same components plus a component representing the change in the terms of trade. An artificial data set is used to illustrate the decomposition of the Moorsteen-Bjurek TFP index.

1. INTRODUCTION

Effective economic and business policy-making requires the accurate measurement of total factor productivity (TFP) change and its components. Economists use numerous measures for this purpose, including Laspeyres, Fisher, Törnqvist, Bennet and Malmquist TFP indexes, and measures of cost, technical, allocative and scale efficiency, to name just a few. Added to this list are the Solow (1957) growth accounting models frequently used to decompose economic growth into measures of input growth and technical change. Empirical researchers sometimes have difficulty choosing one concept or measure over another, and it is not helpful that the relationships between many of them are unclear. One of the aims of this paper is to present these numerous efficiency and productivity concepts, models and measures within a coherent unifying framework.

The framework I propose is both conceptually and mathematically simple. This simplicity is achieved by defining index numbers in terms of aggregate quantities and prices. The idea of defining index numbers in terms of aggregates has been around for a long time. More than twenty-five years ago, for example, Caves, Christensen and Diewert (1982b, p.73) recognised that "a key development in the economic theory of index numbers has been the demonstration that numerous index number formulas can be explicitly derived from particular aggregator functions". Surprisingly, they did not derive their Malmquist TFP index in this way. Except in restrictive

[1] This paper was completed while the author was visiting the Universitat Autònoma de Barcelona. The project was financially supported by the Generalitat de Catalunya. Helpful comments were provided by Marcel Timmer, Bert Balk, Knox Lovell, Prasada Rao, Antonio Perache and Emili Grifell-Tatje.

special cases, their TFP index fails to satisfy a sufficient condition for an unambiguous decomposition of productivity change, namely that it can be written as a function of aggregate quantities and prices.

There are essentially two main approaches to decomposing TFP growth. In the bottom-up approach, researchers define generic measures of efficiency and technical change and then combine them to form a TFP index – see, for example, Balk (2001). In the top-down approach, they start with a recognizable TFP index and then attempt to decompose it in a meaningful way – see, for example, Fare, Grosskopf, Norris and Zhang (1994), Ray and Mukherjee (1996) and Kuosmanen and Sipiläinen (2009). This paper combines the main features of both approaches. I start with input and output aggregator functions that are consistent with axioms from index number theory, then build up to measures of efficiency and technical change, and eventually to recognizable TFP indexes. Such TFP index numbers are said to be *complete*. Because of the manner of their construction, they can be easily decomposed into meaningful measures of technical change and efficiency change. The class of complete TFP indexes includes Fisher and Törnqvist indexes, but not the Malmquist index of Caves, et al. (1982a). TFP indicators of the type proposed by Bennet (1920) are also complete.

The paper is divided into several sections. Section 2 is used to show how Fisher, Törnqvist, Konus and Malmquist price and quantity indexes can be motivated without resort to assumptions concerning either the level of competition in product markets or the returns to scale properties of the production technology. All that is assumed is that non-negative input and output price vectors exist[2]. Section 3 is used to develop a simple two-dimensional geometric representation of TFP and profitability for a multiple-input multiple-output firm. This section is also used to formally define the terms *multiplicative completeness* and *additive completeness* in the context of TFP indexes. Completeness is a crucial requirement for an economically-meaningful decomposition of TFP change. The fact that the Malmquist index of Caves, et al. (1982a) is not complete implies that, except in special cases, it is a biased measure of TFP change – see Grifell-Tatje and Lovell (1995). A further implication is that the popular Fare, et al. (1994) decomposition of that index generally yields unreliable estimates of technical change and/or efficiency change. Section 4 is used to show that important measures of technical, scale and mix efficiency can be defined in terms of quantity aggregates. Measures of revenue, cost, profit and allocative efficiency can also be defined in terms of these aggregates, but these measures are not strictly necessary for a useful decomposition of TFP or profitability change. Section 5 draws all of these concepts together and shows that any multiplicatively complete TFP index can be decomposed into unambiguous measures of technical change and technical efficiency change, as well as natural measures of mix and scale efficiency change. Profitability change can be decomposed into these same components plus a component measuring the change in the terms of trade. Recently, Kuosmanen and Sipiläinen (2009) have used this relationship between profitability, TFP and the terms of trade to write the Fisher TFP index as a product of technical change, efficiency change and price change "effects". Section 6 uses panel data to illustrate the decomposition of the multiplicatively-complete Moorsteen-Bjurek TFP index. The paper is concluded in Section 7.

The contributions of the paper are four-fold. First, it shows how numerous efficiency and productivity concepts can be represented in a coherent aggregate quantity-price framework. Second, it shows that there exists a large class of well-known TFP index numbers that can be decomposed into measures of technical change and efficiency change. These decompositions are exact, they do not necessarily involve measures of profitability change or changes in the terms of trade, and all components except the technical change component can be unambiguously interpreted as measures of efficiency change. Until now, only the Malmquist TFP index of Caves, et al. (1982a) has been decomposed in this way[3]. Third, the paper demonstrates that popular decompositions of the Malmquist TFP index are generally unreliable, if only because the Malmquist index itself is an incomplete

[2] I define and decompose TFP change in a way that is valid even when goods are not priced. However, prices are necessary for the computation of most TFP indexes (e.g., Paasche, Laspeyres, Fisher, Tornquist, Konus) and for the definition and decomposition of profitability change.

[3] The Kuosmanen and Sipiläinen (2009) decomposition of the Fisher TFP index involves a measure of price change, but this price "effect" is not a measure of efficiency change. As this paper demonstrates, measures of price change are not necessary for the decomposition of a concept that is defined in terms of quantities only. The Ray and Mukherjee (1996) decomposition of the Fisher index also includes a scale effect that is not an efficiency change effect.

measure of TFP change. Finally, the paper demonstrates that the Moorsteen-Bjurek index can be decomposed in a way that Balk (1998, p.114) had suggested was impossible. More recently, Balk (2005) had described this index as being "decomposition resistant". Lovell (2003) has decomposed the Moorsteen-Bjurek index into measures of technical change and technical efficiency change, as well as scale and mix effects. However, these scale and mix effects do not lie in the unit interval and cannot be interpreted as efficiency effects. More seriously, Balk (2005, p.25) appears to show that the Lovell (2003) decomposition is mathematically incorrect.

2. PRICE AND QUANTITY INDEXES

There are two main approaches to the construction of index numbers – the axiomatic (or test) approach and the functional (or economic-theoretic) approach. In this section I motivate some important formulas associated with each approach, list their desirable and undesirable properties, and identify conditions under which they are equivalent. To save space, attention is focused on indexes that can be expressed as ratios rather than as differences. To avoid repetition, and also because the Divisia and Törnqvist indexes are most easily motivated in a time-series context, most of this material is presented using the language of inter-temporal comparisons. Also to avoid repetition, I only explicitly deal with the construction of input quantity and/or price indexes.

The Axiomatic Approach

Let $x_t \in \Re_+^N$ and $w_t \in \Re_+^N$ denote vectors of observed input quantities and prices in period *t*. Also let $X_{0t} \equiv X(x_t, x_0)$ and $W_{0t} \equiv W(w_t, w_0)$ denote indexes that measure changes in input quantities and prices between periods 0 and *t* using period 0 as a base. Different formulas are often assessed in terms of whether they satisfy certain axioms or tests. In the case of indexes that can be expressed as ratios, these include[4]:

A.1 Monotonicity axiom[5]: $X(x_m, x_0) > X(x_t, x_0)$ if $x_m \geq x_t$ and $X(x_t, x_m) < X(x_t, x_0)$ if $x_m \geq x_0$.
A.2 Linear homogeneity axiom: $X(\lambda x_t, x_0) = \lambda X(x_t, x_0)$ and $W(\lambda w_t, w_0) = \lambda W(w_t, w_0)$ for $\lambda > 0$.
A.3 Identity axiom: $X(x_t, x_t) = W(w_t, w_t) = 1$.
A.4 Homogeneity of degree 0 axiom: $X(\lambda x_t, \lambda x_0) = X(x_t, x_0)$ and $W(\lambda w_t, \lambda w_0) = W(w_t, w_0)$ for $\lambda > 0$.
A.6 Proportionality axiom: $X(\lambda x_0, x_0) = \lambda$ and $W(\lambda w_0, w_0) = \lambda$ for $\lambda > 0$.

T.1 Transitivity (or circular) test: $X_{0t} = X_{0i} X_{it}$; $W_{0t} = W_{0i} W_{it}$.
T.2 Time reversal test: $X_{0t} = 1/X_{t0}$; $W_{0t} = 1/W_{t0}$.
T.3 Product test: $C_{0t} = W_{0t} X_{0t}$.
T.4 Factor reversal test: $C_{0t} = W_{0t} X_{0t}$ where W_{0t} and X_{0t} have the same functional form.
T.6 Consistency in aggregation test: an aggregate index $X(x_t, x_0)$ can also be constructed as an index of sub-aggregates, where all indexes have the same functional form.
T.7 Equality test: if indexes of sub-aggregates are all equal to λ then $X(x_t, x_0) = \lambda$.

where $C_{0t} = C_t / C_0$ is a simple cost index and $C_t = w_t' x_t$ is total cost in period *t*. There are a number of inherent conflicts between these axioms and tests. For example, Balk (1995, pp.76, 86) claims there are no price indexes that simultaneously satisfy A.3, T.1 and T.4, and that the only price-quantity index pairs that satisfy A.2, A.6, T.3, T.6 and T.7 are the Paasche-Laspeyres and Laspeyres-Paasche pairs. Of course, not all axioms and tests are equally important in all empirical contexts. For example, the transitivity test T.1 is usually regarded as being especially important in a cross-section context where there is no natural ordering of the data points; the factor reversal test T.4 is mathematically convenient but is arguably of little practical value if a price-quantity index number pair already satisfies the product test T.3; and the product test T.3 is itself something of an empty test because, given any price (quantity) index number, an implicit quantity (price) index can be trivially defined so that T.3 is satisfied.

---

[4] For more details, see Balk (1995, pp.71-85)

[5] The notation $x_m \geq x_t$ means that $x_{im} \geq x_{it}$ for all $i = 1,...,N$ and there exists at least one value $i \in \{1,...,N\}$ where $x_{im} > x_{it}$.

It is useful to begin a discussion of specific index number formulas by noting that the ratios $w_0'x_t / w_0'x_0$ and $w_t'x_t / w_t'x_0$ only depart from unity as x_t departs from x_0. This property makes them natural measures of input quantity change. They are, in fact, the Laspeyres and Paasche input quantity indexes, and their geometric average is the Fisher[6] input quantity index:

(2.1) $$X_{0t}^F = \left(X_{0t}^P X_{0t}^L \right)^{1/2} \quad \text{where} \qquad \text{(Fisher)}$$

(2.2) $$X_{0t}^L = w_0'x_t / w_0'x_0 \quad \text{and} \qquad \text{(Laspeyres)}$$

(2.3) $$X_{0t}^P = w_t'x_t / w_t'x_0 \qquad \text{(Paasche)}$$

The Fisher input price index W_{0t}^F can be obtained from X_{0t}^F by interchanging the roles of prices and quantities. The Fisher index is "ideal" in the sense it satisfies almost all axioms and tests, including the factor reversal test. However, it does not satisfy the transitivity (circularity) test.

If prices and quantities are measured at discrete points in time, the percentage change in costs between periods 0 and t is $(C_t - C_0)/C_0$. It is well-known that the Paasche and Laspeyres indexes can be derived from two alternative decompositions of this percentage change (see Appendix A). On the other hand, if prices and quantities are viewed as continuous functions of time, the percentage change in costs is $\dot{C}/C_t$. Divisia input quantity and price indexes can be derived from a decomposition of this percentage change (see Appendix B) without imposing any restrictions on market behaviour or the structure of the production technology. In particular, input markets do not need to be competitive, firms do not need to minimise costs, technical change does not need to be Hicks-neutral, and the technology does not need to exhibit constant returns to scale[7]. The practical usefulness of the Divisia index is nevertheless limited by the fact that it is a continuous-time index. Discrete-time approximations are available, although they can only be viewed as such if i) they are used in a time-series context, and ii) they are used to make comparisons between quantities in adjacent periods of time. Perhaps the best-known approximation is the Törnqvist index:

(2.4) $$X_{t-1,t}^T = \exp\left\{ \sum_{n=1}^{N} \left(\frac{s_{nt} + s_{n,t-1}}{2} \right) \left(\ln x_{nt} - \ln x_{n,t-1} \right) \right\} \qquad \text{(Törnqvist)}$$

where $s_{nt} = w_{nt} x_{nt} / C_t$ is the n-th input cost share in period t. To make comparisons between periods that are not adjacent, the Törnqvist index can be chained:

(2.5) $$X_{0t}^T = X_{01}^T X_{12}^T \ldots X_{t-1,t}^T . \qquad \text{(chained Törnqvist)}$$

---

[6] The idea of using the geometric mean of the Paasche and Laspeyres indexes can be traced back at least as far as Pigou (1912). However, it was Fisher (1922) who examined the properties of the resulting index and deemed it (almost) ideal.

[7] This is contrary to some claims made in the literature - see, for example, Nadiri and Nandi (1999, p.488). The confusion may be due to the fact that some restrictions are required for the Divisia index to satisfy certain invariance axioms, including, for example, path invariance. Path invariance means that "the index at time *t* depends on the [input] levels at time *t*, and not on the historical route by which the levels were attained" – Richter (1966, p.751). It can be seen from the definitions in Appendix B that Divisia aggregate prices and quantities are line integrals – they are areas under functions that describe the time paths of prices and quantities. Thus, they will not generally be path invariant. Hulten (1973) establishes sufficient conditions under which a Divisia quantity index is path invariant, and one of these conditions is that the technology exhibits constant returns to scale. If the paths are sectionally smooth, then constant returns to scale is also a necessary condition for path invariance. Whether or not any of these types of invariance axioms are desirable or not is another matter. For example, Richter (1966) seeks output indexes that are invariant to movements along the boundary of a production possibilities frontier – after all, he says, "we are only changing the mix of outputs, so it is not clear why an output index [or productivity] should change" - Richter (1966, pp.742-743). Nowadays, such movements are regarded as important components of measured productivity change – see, for example, Balk (2005, p.2).

The chained Törnqvist direct input price index W_{0t}^{T} is obtained from X_{0t}^{T} by interchanging the roles of prices and quantities. The Divisia index satisfies the factor reversal test (see Appendix B) but the Törnqvist index does not. Like the Fisher index, the Törnqvist index fails the transitivity test.

The Functional Approach

The functional approach to index number construction exploits various alternative representations of the production technology, including the cost, revenue, profit and input and output distance functions:

(2.6) $$C^k(w_t, q_t) = \min_{x_t \geq 0}\{w_t' x_t : (x_t, q_t) \in T^k\}$$ (cost function)

(2.7) $$R^k(p_t, x_t) = \max_{q_t \geq 0}\{p_t' q_t : (x_t, q_t) \in T^k\}$$ (revenue function)

(2.8) $$\pi^k(p_t, w_t) = \max_{x_t \geq 0, q_t \geq 0}\{p_t' q_t - w_t' x_t : (x_t, q_t) \in T^k\}$$ (profit function)

(2.9) $$D_I^k(x_t, q_t) = \max_{\rho}\{\rho > 0 : (x_t / \rho, q_t) \in T^k\}$$ (input distance function)

(2.10) $$D_O^k(x_t, q_t) = \min_{\delta}\{\delta > 0 : (x_t, q_t / \delta) \in T^k\}$$ (output distance function)

where $q_t \in \Re_+^M$ and $p_t \in \Re_+^M$ are vectors of output quantities and prices, and T^k denotes the period-k production possibilities set. The cost function gives the minimum cost of producing the output vector given input prices; the revenue function gives the maximum revenue that can be produced using an input vector and given output prices; the profit function gives the maximum profit that can be achieved given input and output prices; the input distance function gives the maximum factor by which a firm can radially contract its input vector and still produce the same outputs; and the output distance function gives the inverse of the largest factor by which a firm can radially expand its output vector while using the same inputs. The duality relationships between these functions have been established by Shephard (1953, 1970) and Fare and Primont (1994). If the production technology is regular then these functions possess a number of important properties. With a view to eventually decomposing TFP indexes, perhaps the most important of these properties are that the cost and revenue functions are non-negative and linearly homogeneous and non-decreasing in prices, while the output and input distance functions are non-negative and linearly homogeneous and non-decreasing in output quantities and input quantities respectively.

To motivate cost-based index formulas, observe that the ratios $C^t(w_t, q_t)/C^t(w_0, q_t)$ and $C^0(w_t, q_0)/C^0(w_0, q_0)$ only depart from unity as w_t departs from w_0, making them natural measures of input price change[8]. Indeed, they are isomorphic to the true cost of living index described by Konus (1924) and discussed in a survey article by Frisch (1936, pp.10-13). What I refer to in this paper as the Konus input price index is the geometric average of these two ratios:

(2.11) $$W_{0t}^{K} = \left(\frac{C^t(w_t, q_t) C^0(w_t, q_0)}{C^t(w_0, q_t) C^0(w_0, q_0)} \right)^{1/2}$$ (Konus)

The associated (implicit) input quantity index is $X_{0t}^{K} = C_{0t} / W_{0t}^{K}$. Thus, by construction, W_{0t}^{K} and X_{0t}^{K} satisfy the product test. According to Balk (1998), the Konus input price index has a number of theoretically and intuitively appealing properties, but it does not satisfy the transitivity or factor reversal tests, and cannot be computed without knowing something about the cost function. It is common to either i) estimate the cost function using parametric or non-parametric techniques, or ii) make use of specific functional form and behavioural assumptions to express the index in terms of observable quantities, such as prices and cost shares.

To motivate distance-based formulas, observe that the ratios $D_I^t(x_t, q_t)/D_I^t(x_0, q_t)$ and $D_I^0(x_t, q_0)/D_I^0(x_0, q_0)$ only depart from unity as x_t departs from x_0, making them natural measures of input quantity change. Caves,

[8] Simpler versions of these ratio's are available in the restrictive special case where the technology exhibits input homotheticity. In that case, both ratios are independent of output quantities – see, for example, Balk (1998).

et al. (1982a) note that they are the producer analogues of a distance-based index measure first suggested in a consumer context by Malmquist (1953). The Malmquist input quantity index is the geometric average of these two ratios:

$$X_{0t}^{M} = \left(\frac{D_I^t(x_t, q_t) D_I^0(x_t, q_0)}{D_I^t(x_0, q_t) D_I^0(x_0, q_0)} \right)^{1/2} \tag{2.12}$$

(Malmquist)

This input quantity index can, of course, be used to obtain an implicit input price index: $W_{0t}^{M} = C_{0t} / X_{0t}^{M}$. Again, by construction, these indexes satisfy the product test However, their theoretical appeal is again overshadowed by the fact that they do not satisfy the transitivity or factor reversal tests, and they cannot be computed without knowing something about the production technology. Again, it is common to either i) estimate the period 0 and period *t* distance functions using parametric or non-parametric techniques, or ii) make use of specific functional form and behavioural assumptions to express them in terms of observable quantities.

Exact and Superlative Index Numbers

Earlier in this section, the axiomatic approach to index number construction was used to motivate the use of Törnqvist indexes for intertemporal comparisons. However, Törnqvist indexes are also frequently used in a multilateral (cross-section) context where, for example, the index number for comparing the input levels of firms 0 and *t* takes the form:

$$X_{0t}^{T} = \exp\left\{ \sum_{n=1}^{N} \left(\frac{s_{nt} + s_{n0}}{2} \right) \left(\ln x_{nt} - \ln x_{n0} \right) \right\} \tag{2.13}$$

(Törnqvist)

One rationale for using the Törnqvist formula for multilateral comparisons is provided by Caves, et al. (1982a, p.1398). These authors establish that if i) firms use positive amounts of all inputs, ii) firms are cost minimizers, and iii) the input distance functions $D_I^t(.)$ and $D_I^0(.)$ are translog functions with identical second-order coefficients, then the Törnqvist input quantity index given by (2.13) is equal to the Malmquist input quantity index given by equation (2.12). They also establish an analogous result for output quantity indexes, namely that if i) firms produce positive amounts of all outputs, ii) firms are revenue maximisers, and iii) the output distance functions $D_O^t(.)$ and $D_O^0(.)$ are translog functions with identical second-order coefficients, then the Törnqvist output quantity index is equal to the Malmquist output quantity index. To use the terminology of Diewert (1976), the Törnqvist index is *exact* for a translog functional form. Moreover, because the translog is a second-order flexible functional form, the Törnqvist index is *superlative*.

These results are of enormous practical value insofar as they provide conditions under which theoretically-appealing Malmquist indexes can be computed without having to estimate the parameters of the production technology. However, the positivity and functional form conditions are restrictive. The positivity conditions may be especially restrictive in industries such as agriculture where firms often rationally choose not to use some inputs in the production process (e.g., pesticides), and where outputs are often zero. The functional form conditions are restrictive insofar as the output and input distance functions cannot both be translog unless the technology exhibits constant returns to scale. In that case, there is no unique input-output combination that both maximises revenue and minimises cost. One implication for TFP measurement is that the ratio of Törnqvist output and input indexes cannot be exact for the ratio of Malmquist output and input indexes unless the technology exhibits constant returns to scale – see Caves, et al. (1982a, pp.1404,1407).

Analogous results are available for other output and input quantity and price indexes. For example, Caves, et al. (1982a) establish that if i) firms are cost minimizers, and ii) the cost functions $C^t(.)$ and $C^0(.)$ are translog functions with identical second-order coefficients, then the Törnqvist input price index is equal to the Konus input price index. This result is more powerful than the Malmquist result because it does not require positive amounts of all inputs. As a final example, Diewert (1976) shows that the Fisher index is superlative for a quadratic mean of order two functional form[9].

3. PRODUCTIVITY AND PROFITABILITY INDEXES

Any of the foregoing index number formulas can be used to build intertemporal or multilateral indexes of productivity and profitability change. In this section, I persist with the vocabulary of cross-sectional comparisons and consider the problem of comparing the productivity and profitability of two firms: a reference firm that faces prices (w_0, p_0) and selects the input-output combination (x_0, q_0) from the production possibilities set T^0, and a comparison firm that faces prices (w_t, p_t) and selects (x_t, q_t) from the production possibilities set T^t. Any results that are peculiar to intertemporal comparisons are relegated to footnotes.

Total Factor Productivity Indexes

For single-input single-output firms, total factor productivity (TFP) is almost always defined as the output-input ratio. It is also possible to define TFP as the output minus the input, but this is much less common, partly because it yields an index that is sensitive to units of measurement. Both measures can be generalized to the multiple-output multiple-input case. For example, the TFP of a multiple-output multiple-input firm is commonly defined as the ratio of an aggregate output to an aggregate input – see Jorgenson and Griliches (1967, p.252, eq.253) and Good, Nadiri and Sickles (1997, p.17-19). This is the definition of TFP used throughout this section.

Let $Q_t \equiv Q(q_t)$ and $X_t \equiv X(x_t)$ denote the scalar aggregate output and input associated with the vectors x_t and q_t. Then the TFP of firm t is simply $TFP_t = Q_t / X_t$. The associated TFP index number that measures the change in TFP between firms 0 and t is $TFP_{0t} = TFP_t / TFP_0 = Q_{0t} / X_{0t}$ where $Q_{0t} = Q_t / Q_0$ and $X_{0t} = X_t / X_0$ are indexes measuring changes in aggregate outputs and inputs[10]. Clearly, if inputs are held fixed then $X_{0t} = 1$ and TFP_{0t} is a measure of output change. If firms also happen to be fully efficient in production then TFP_{0t} corresponds to the Caves, et al. (1982a, p.1401) concept of an output-based productivity index; if outputs are held fixed and firms are fully efficient then TFP_{0t} corresponds to their concept of an input-based productivity index.

According to Balk (2003), the idea of measuring TFP change as the ratio of an output quantity index to an input quantity index can be traced back at least as far as Copeland (1937). Figure 1 depicts this long-standing measure of TFP change in aggregate quantity space. In this figure, the TFP for firm t is given by the slope of the ray passing through the origin and point A, while the TFP for firm 0 is given by the slope of the ray passing through the origin and point Z. Let a and z denote the angles between the horizontal axis and the rays passing through points A and Z. Then the change in TFP between firms 0 and t can be conveniently written $TFP_{0t} = \text{slope 0A} / \text{slope 0Z} = \tan a / \tan z$. Being able to write a TFP index as the ratio of (tangent) functions of angles in aggregate quantity space is particularly useful for conceptualising alternative decompositions of TFP change. To illustrate, let e denote the angle between the horizontal axis and the ray passing through the origin and *any* non-negative point E in aggregate quantity space. Then the change in TFP between firms 0 and t can be decomposed as $TFP_{0t} = \tan a / \tan z = (\tan a / \tan e)(\tan e / \tan z)$. Within this framework, an infinite number of points E can be used to effect a decomposition of TFP change. Section 4 below will focus on points that feature in the empirical measurement of efficiency and technical change.

---

[9] Diewert attributes the proof to Byushgens (1925), Konus and Byushgens (1926), Frisch (1936, p.30), Wald (1939, p.331), Afriat (1972, p.45) and Pollack (1971).

[10] In the continuous-time case, the growth rate in TFP is usually defined as the difference between the growth rate of outputs and the growth rate of inputs: $\dot{TFP} / TFP = \dot{Q} / Q - \dot{X} / X$. See, for example, Jorgenson and Griliches (1967, p.252, eq.254). To see the relationship between the discrete and continuous time definitions of TFP growth/change, first note that if the variable Z is a continuous function of time then the rate of growth in Z is $\dot{Z} / Z = d \ln Z / dt$. Two alternative discrete-time approximations to this growth rate are $\Delta Z_t / Z_{t-1} = (Z_t - Z_{t-1}) / Z_{t-1}$ and $\Delta \ln Z_t = \ln Z_t - \ln Z_{t-1}$. In a cross-section context, corresponding measures of the difference between Z_t and Z_0 are $(Z_t - Z_0) / Z_0 = Z_{0t} - 1$ and $\ln Z_t - \ln Z_0 = \ln Z_{0t}$ where $Z_{0t} \equiv Z_t / Z_0$.

In practice, computing a TFP index implicitly involves choosing output and input quantity aggregator functions $Q(.)$ and $X(.)$ from a limited range of functional forms. The menu of functional forms is limited by the requirement that the indexes $Q_{0t} = Q_t / Q_0$ and $X_{0t} = X_t / X_0$ should satisfy as many as possible of the axioms and tests discussed in Section 2. Axioms A.1 to A.6 limit the class of admissible input quantity aggregator functions to those that are non-negative and non-decreasing and linearly homogeneous in inputs. If we have information on prices, quantities and the production technologies used by firms 0 and *t* then the following are just four of the non-negative non-decreasing linearly-homogenous aggregator functions that are available for aggregating the inputs of firm $k \in \{0,t\}$:

(3.1) $X(x_k) = w_0' x_k$ (Laspeyres)

(3.2) $X(x_k) = w_t' x_k$ (Paasche)

(3.3) $X(x_k) = D_I^0(x_k, q_0)$ (Malmquist)

(3.4) $X(x_k) = D_I^t(x_k, q_t)$ (Malmquist)

The intuition behind the linear aggregator functions (3.1) and (3.2) is that we should take a weighted sum of individual inputs, using weights that reflect the relative importance, or value, of each input to the firm. The aggregator function (3.1) uses the prices paid by firm 0 as weights, while the aggregator function (3.2) uses prices paid by firm *t* as weights. The aggregator functions (3.3) and (3.4) can be viewed as non-linear aggregator functions that are rooted in the available technology and use firm 0 outputs and firm *t* outputs as weights. The practical disadvantage of these distance aggregator functions is that they must be estimated. Other non-negative linearly-homogeneous functions that must also be estimated but have nevertheless been used to implicitly aggregate input prices and quantities for firm $k \in \{0,t\}$ include:

(3.5) $W(w_k) = C^0(w_k, q_0)$ (Konus)

(3.6) $W(w_k) = C^t(w_k, q_t)$ (Konus)

(3.7) $X(x_k) \propto \exp\left\{\sum_{n=1}^{N}\left(\frac{s_{nt}+s_{n0}}{2}\right)(\ln x_{nt} - \ln x_{n0})\right\}$ (Törnqvist)

(3.8) $X(x_k) = \left(\sum_{n=1}^{N}\sum_{h=1}^{N}\beta_{nh} x_{nk}^{r/2} x_{hk}^{r/2}\right)^{1/r}$ (quadratic mean)

(3.9) $X(x_k) = \left(\prod_{n=1}^{N} x_{nk}^{\beta_n}\right)$ where $\sum_{n=1}^{N}\beta_n = 1.$ (Cobb-Douglas)

The aggregator functions given by (3.1) and (3.2) are referred to as Laspeyres and Paasche aggregator functions because if the input vectors of firms 0 and *t* are both aggregated using these function then the resulting input quantity indexes are respectively the Laspeyres and Paasche indexes defined by equations (2.2) and (2.3). Not surprisingly, the geometric mean of the Laspeyres and Paasche aggregator functions is itself a non-negative non-decreasing and linearly-homogeneous aggregator function that yields the Fisher index. The aggregator functions (3.3) and (3.4) are referred to as Malmquist aggregator functions because they yield the firm 0 and firm *t* Malmquist input indexes of Caves, et al. (1982a, p.1396). If inputs are aggregated using the geometric mean of (3.3) and (3.4) then we obtain the Malmquist input quantity index defined in equation (2.12). The Konus aggregator functions (3.5) and (3.6) lead to the Konus input price index defined in (2.11), while the Törnqvist aggregator function (3.7) leads to the multilateral Törnqvist index given by (2.13). Of course, the Törnqvist index is only theoretically meaningful in a cross-section context if the functional form and positivity restrictions of Caves, et al. (1982a) hold[11]. Good, et al. (1997, p.24) observe that the quadratic mean of order *r* aggregator

[11] In a time-series context, the following aggregator function is available:

function (3.8) nests the generalized Leontief (r = 1), quadratic (r = 2), translog $(r \to 0)$, constant elasticity of substitution ($\beta_{nh} = 0$ for all $n \neq h$) and Cobb-Douglas ($\beta_{nh} = 0$ for all $n \neq h$ and $r \to 0$) functional forms as special cases. The constant returns to scale Cobb-Douglas aggregator function (3.9) was used by Solow (1957) in his seminal paper on decomposing economic growth. The Solow (1957) model is discussed in more detail in Section 5.

The foregoing demonstrates that ratio-type output and input indexes are underpinned by specific output and input aggregator functions. In turn, TFP indexes can be formed by taking ratios of these output and input indexes. Although it is possible to construct a TFP index by, for example, dividing a Fisher output quantity index by a Törnqvist input quantity index, it is much more common[12] to select output and input quantity indexes that have the same functional form. In keeping with that tradition, the Fisher, Törnqvist and Konus TFP indexes, for example, are

$$(3.10) \qquad TFP_{0t}^{F} = \frac{Q_{0t}^{F}}{X_{0t}^{F}} = \left(\frac{p_t'q_t}{p_t'q_0}\frac{p_0'q_t}{p_0'q_0}\right)^{1/2}\left(\frac{w_t'x_0}{w_t'x_t}\frac{w_0'x_0}{w_0'x_t}\right)^{1/2} \qquad \text{(Fisher)}$$

$$(3.11) \qquad TFP_{0t}^{T} = \frac{Q_{0t}^{T}}{X_{0t}^{T}} = \exp\left\{\sum_{m=1}^{M}\left(\frac{r_{mt}+r_{m0}}{2}\right)(\ln q_{mt} - \ln q_{m0}) - \sum_{n=1}^{N}\left(\frac{s_{nt}+s_{n0}}{2}\right)(\ln x_{nt} - \ln x_{n0})\right\} \qquad \text{(Törnqvist)}$$

$$(3.12) \qquad TFP_{0t}^{K} = \frac{Q_{0t}^{K}}{X_{0t}^{K}} = \rho_{0t}\left(\frac{R^t(p_0,x_t)R^0(p_0,x_0)}{R^t(p_t,x_t)R^0(p_t,x_0)}\right)^{1/2}\left(\frac{C^t(w_t,q_t)C^0(w_t,q_0)}{C^t(w_0,q_t)C^0(w_0,q_0)}\right)^{1/2} \qquad \text{(Konus)}$$

where $r_{mt} = p_{mt}q_{mt}/R_t$ is the m-th revenue share for firm t and ρ_{0t} is an index of profitability change formally defined in equation (3.16) below. Perhaps surprisingly, taking the ratio of Malmquist output and input quantity indexes does not result in the Malmquist productivity indexes of Caves, et al. (1982a). Rather the ratio of the Malmquist output index to the Malmquist input index given by (2.12) results in an index suggested by Moorsteen (1961) and Bjurek (1996)[13]:

$$(3.13) \qquad TFP_{0t}^{BM} = \frac{Q_{0t}^{M}}{X_{0t}^{M}} = \left(\frac{D_O^t(x_t,q_t)D_O^0(x_0,q_t)}{D_O^t(x_t,q_0)D_O^0(x_0,q_0)}\right)^{1/2}\left(\frac{D_I^t(x_0,q_t)D_I^0(x_0,q_0)}{D_I^t(x_t,q_t)D_I^0(x_t,q_0)}\right)^{1/2} \qquad \text{(Moorsteen-Bjurek)}$$

Difference-type output and input indexes of the type proposed by Bennet (1920) are also underpinned by specific output and input aggregator functions, including, for example, the directional distance (or shortage) function of Luenberger (1992). In turn, these difference-type input and output indexes can be used to construct Bennet-type TFP indexes. These difference-type indexes and the ratio-type indexes given by (3.10) to (3.13) satisfy a fundamentally important precondition for a meaningful decomposition of productivity change, namely completeness. Two types of completeness can be defined corresponding to the ratio- and difference-type index frameworks:

$$X(x_t) = X(x_{t-1})\exp\left\{\sum_{n=1}^{N}\left(\frac{s_{nt}+s_{n,t-1}}{2}\right)(\ln x_{nt} - \ln x_{n,t-1})\right\}$$

If this function is used to aggregate input quantities in adjacent periods then we obtain the Tornquist index defined by equation (2.4).

[12] This common practice is not always theoretically desirable. For example, the ratio of Tornquist output and input quantity indexes cannot be regarded as a superlative TFP index in the sense of Diewert (1976) unless the technology exhibits constant returns to scale (see Section 2).

[13] Lovell (2003) refers to the Moorsteen-Bjurek index as the Malmquist total factor productivity index, and to the Caves, et al. (1982a) Malmquist index as the Malmquist productivity index . Bjurek (1996) uses the term Malmquist index in reference to both indexes. Diewert (1992) suggests that the index might also have been suggested by Hicks (1961). I follow Balk (2005) and refer to it as the Moorsteen-Bjurek index.

D.1. Multiplicative Completeness: Let $TFP(x_t, q_t, x_0, q_0)$ denote an index number that measures the difference in TFP between firms/periods 0 and *t* using firm/period 0 as a base. $TFP(x_t, q_t, x_0, q_0)$ is *multiplicatively complete* if and only if it can be expressed in the form

$$TFP(x_t, q_t, x_0, q_0) = \frac{Q(q_t)/X(x_t)}{Q(q_0)/X(x_0)}$$

where $Q(.)$ and $X(.)$ are non-negative non-decreasing linearly-homogeneous functions.

D.2. Additive Completeness: Let $TFP(x_t, q_t, x_0, q_0)$ denote an index number that measures the difference in TFP between firms/periods 0 and *t* using firm/period 0 as a base. $TFP(x_t, q_t, x_0, q_0)$ is additively *complete* if and only if it can be expressed in the form

$$TFP(x_t, q_t, x_0, q_0) = Q(q_t) - Q(q_0) - X(x_t) + X(x_0)$$

where $Q(.)$ and $X(.)$ are non-negative non-decreasing functions satisfying the translation property $Q(\lambda q) = \lambda + Q(q)$ and $X(\lambda x) = \lambda + X(x)$.

Multiplicatively complete TFP indexes are theoretically plausible insofar as they satisfy the index number axioms A.1 to A.6. Additively complete TFP indexes satisfy a similar set of axioms. Importantly, completeness is a sufficient condition for decomposing a TFP index to into measures of technical change, technical efficiency change, scale efficiency change and mix efficiency change. In this paper I provide details for multiplicatively complete TFP indexes.

Even though Caves, et al. (1982a) proposed the use of Malmquist output and input quantity indexes of the form given by equation (2.12), they did not use ratios of those indexes to construct a complete TFP index – that idea was first taken up seriously by Bjurek (1996). Instead, Caves, et al. (1982a) defined indexes that are complete if and only if the technology is of a restrictive form. The most popular forms of the Caves, et al. (1982a) Malmquist indexes are[14]:

(3.14) $$TFP_{0t}^{OM} = \left(\frac{D_O^t(x_t, q_t) D_O^0(x_t, q_t)}{D_O^t(x_0, q_0) D_O^0(x_0, q_0)} \right)^{1/2}$$ (output-oriented Malmquist)

(3.15) $$TFP_{0t}^{IM} = \left(\frac{D_I^t(x_0, q_0) D_I^0(x_0, q_0)}{D_I^t(x_t, q_t) D_I^0(x_t, q_t)} \right)^{1/2}$$ (input-oriented Malmquist)

Fare, Grosskopf and Roos (1998, p.136) show that the output-oriented Malmquist index given by (3.14) equals the Moorsteen-Bjurek index given by (3.13) if and only if the technology is inversely homothetic and exhibits constant returns to scale[15]. Unless these conditions hold, the Malmquist index is a systematically biased measure of changes in productivity – see Grifell-Tatje and Lovell (1995).

Profitability Indexes

Associated with the output and input quantity aggregates discussed above are price aggregates $P_t \equiv P(p_t)$ and $W_t \equiv W(w_t)$. Not only must the aggregator functions $P(.)$ and $W(.)$ be non-negative and non-decreasing and linearly homogeneous in prices, they must be chosen in such a way that any quantity-price aggregator function pairs satisfy the product rules $P_t Q_t = p_t' q_t$ and $W_t X_t = w_t' x_t$. Eichhorn (1978, p.144) deomonstrates that if the

[14] See, for example, Fare, Grosskopf, Lindgren and Roos (1992, p.90) and Lovell (2003, p.440).

[15] Fare, et al. (1998) attribute the proof to Fare, Grosskopf and Roos (1996). A technology is said to exhibit inverse homotheticity if $D_O(x,q) = D_O(a,q)/F(D_I(x,a))$ where *F* is increasing and a is an arbitrary vector. If the technology exhibits constant returns to scale then *F* is the identity function.

price aggregator functions $P(.)$ and $W(.)$ depend only on prices and not on quantities, and if the quantity aggregator functions $Q(.)$ and $X(.)$ depend only on quantities and not on prices, there are no price-quantity aggregator function pairs that satisfy the product rules. However, the product rules can be easily satisfied if both prices and quantities are permitted to enter each aggregator function. Indeed, the non-negativity, monotonicity, linear homogeneity and product rule requirements can all be trivially satisfied by choosing $P_t = p_t'q_t / Q_t$ and $W_t = w_t'x_t / X_t$. Henceforth, I will assume all aggregate prices and quantities are chosen to satisfy the product rules.

A common measure of business performance is profitability, defined as the ratio of revenue to cost: $\rho_t = R_t / C_t$ where $R_t \equiv p_t'q_t$ and $C_t \equiv w_t'x_t$. The associated index number that measures the change in profitability between firms 0 and t is $\rho_{0t} = \rho_t / \rho_0 = R_{0t} / C_{0t}$ where $R_{0t} = R_t / R_0$ and $C_{0t} = C_t / C_0$ are simple revenue and cost indexes. If the product rules are satisfied then

(3.16) $$\rho_{0t} = \frac{R_{0t}}{C_{0t}} = \frac{P_{0t}Q_{0t}}{W_{0t}X_{0t}} = TFP_{0t} \times \frac{P_{0t}}{W_{0t}} \qquad (\Delta\text{profitability} = \Delta\text{TFP} \times \Delta\text{terms of trade})$$

Thus, the change in profitability between firms 0 and t can be decomposed into the product of indexes measuring the change in TFP and the change in the terms of trade[16]. In the productivity literature it is not uncommon to re-arrange (3.16) and compute an "indirect" TFP index as a deflated revenue index divided by a deflated cost index[17]:

(3.17) $$TFP_{0t} = \frac{R_{0t} / P_{0t}}{C_{0t} / W_{0t}}.$$

It is clear that if profitability is unchanged (e.g., if input and output markets are both perfectly competitive) then a TFP index can be computed as the inverse of the change in the terms of trade: $TFP_{0t} = W_{0t} / P_{0t}$. On the other hand, if the terms of trade are unchanged (e.g., in a cross-section context where all firms face identical prices) then $P_{0t} = W_{0t}$ and TFP change is equal to profitability change[18]: $TFP_{0t} = \rho_{0t}$.

Other monetary measures of business performance are available, including profit: $\pi_t = R_t - C_t$. Grifell-Tatje and Lovell (1999) consider an additively complete TFP index and show how profit change can be (additively) decomposed into six components representing technical efficiency change, technical change, scale change, resource-mix change, product-mix change, and price change. Sahoo and Tone (2008) implement a variant of this decomposition using non-radial data envelopment analysis (DEA). In this paper I focus on the profitability measure, partly because it lends itself to a very simple multiplicative decomposition,[19] and partly because it is invariant to units of measurement. These two properties are illustrated geometrically in Figure 2. This figure shows that profitability for firm t can be decomposed into the product of a tangent function of the angle a (measuring TFP) and a tangent function of the angle j (measuring the terms of trade). Profitability is invariant to units of measurement as evidenced by the fact it does not depend on the lengths of the rays passing through the quantity and price points A and J – it only depends on the angles between those rays and the horizontal axis. In contrast, profit has the dimension of money because it *is* a function of the lengths of those rays – it is proportional to the length of the ray passing through A, where the factor of proportionality depends on the length of the ray passing through J. Profit is only invariant to units of measurement in the special case where the price and

---

[16] In continuous time, the growth rate in profitability can be written as the growth rate in TFP plus the difference between the growth rates of output and input prices: $\dot{\rho} / \rho = \dot{TFP} / TFP + \dot{P} / P - \dot{W} / W$.

[17] See, for example, Balk (2003, p.19).

[18] Note that profitability can be unchanged even when markets are uncompetitive, and terms of trade can be unchanged even when firms face different prices. Furthermore, if both input and output markets are perfectly competitive and all firms face identical prices, then $TFP_{st} = 1$.

[19] Aside from simplicity, multiplicative decompositions are attractive because they are preserved under an EKS (transitivity) transformation, which, in the present context, means that $\rho_{0t}^{EKS} = TFP_{0t}^{EKS} \times P_{0t}^{EKS} / W_{0t}^{EKS}$.

quantity vectors are orthogonal, in which case profits are zero[20].

4. EFFICIENCY CONCEPTS

As a precursor to decomposing multiplicatively complete TFP indexes into various components, it is necessary to define conventional efficiency concepts in terms of some of the output and input aggregates introduced in Section 3. In a multiplicative world, efficiency is defined as the (inverse of the) ratio of an observed quantity to some maximum or minimum possible. For example, cost efficiency is the inverse of the ratio of observed cost to the minimum cost possible holding input prices and output quantities fixed. In this section, I define ratio measures of technical, scale and mix efficiency for a firm that selects the input-output combination (x_t, q_t) from the period-*t* production possibilities set. Technical and pure scale efficiency measures are defined in terms of technically-feasible input and output vectors that can be written as scalar multiples of x_t and q_t, which is to say the input and output mixes are held fixed. Mix efficiency measures are then defined in terms of input and output vectors that are technically feasible when the input and output mixes are free to vary.

Technical Efficiency

Since the work of Farrell (1957), the most-widely accepted measure of *output-oriented technical efficiency* (OTE) has been the ratio of observed aggregate output to the maximum aggregate output possible while holding the input vector and the output mix fixed[21]. Similarly, the conventional measure of *input-oriented technical efficiency* (ITE) can be defined as the inverse of the ratio of the observed aggregate input to the minimum aggregate input possible holding the output vector and the input mix fixed. The scalar aggregates associated with the vectors x_t and q_t are $X_t = X(x_t)$ and $Q_t = Q(q_t)$ where $Q(.)$ and $X(.)$ are non-negative linearly homogeneous aggregator functions. Thus, the maximum aggregate output that is technically feasible when using x_t to produce a scalar multiple of q_t is $\bar{Q}_t \equiv Q_t / D_O^t(x_t, q_t)$, while the minimum aggregate input possible when using a scalar multiple of x_t to produce q_t is $\bar{X}_t \equiv X_t / D_I^t(x_t, q_t)$. Thus, Farrell-type measures of technical efficiency can be formally written

$$(4.1) \qquad OTE_t = \frac{Q_t}{\bar{Q}_t} = D_O^t(x_t, q_t) \qquad \text{and} \qquad ITE_t = \frac{\bar{X}_t}{X_t} = D_I^t(x_t, q_t)^{-1}.$$

To represent these concepts graphically, first consider the special case where the firm uses $x_t = (x_{1t}, x_{2t})'$ to produce $q_t = (q_{1t}, q_{2t})'$ and where, without loss of generality, the input and output aggregator functions are both linear[22]: $X(x_t) = \beta_1 x_{1t} + \beta_2 x_{2t}$ and $Q(q_t) = \alpha_1 q_{1t} + \alpha_2 q_{2t}$. Figure 3 depicts this special case in input space. The curved line passing through point B in Figure 3 traces out all technically-efficient input combinations that can produce q_t, while the dashed line passing through point A is an isoinput line that maps out all points that have the same aggregate input as the firm operating at point A (previously referred to as firm *t*, but henceforth also referred to as firm A). If both the input mix and output vector are held fixed then firm A can minimize aggregate input use by radially contracting inputs to point B. Indeed, the ratio of the distance 0B to the distance 0A in Figure 3 is the input-oriented measure of technical efficiency defined in equation (4.1): $ITE_t = \bar{X}_t / X_t = \|B\| \|A\|$. Figure 4 depicts the input-output choice of firm A in output space. In Figure 4, the curved line passing through point C is a production possibilities frontier, while the dashed line passing through point A maps out all points that have the same aggregate output as firm A. Firm A can increase its aggregate output by radially expanding

---

[20] In Euclidean geometry, the length or *norm* of the vector *z* is given by $\|z\| = \sqrt{z'z}$ (Pythagoras's Theorem). For any conformable vectors *y* and *z*, the inner product can be written $y'z = \|y\|\|z\|\cos\theta$ where θ is the angle between *y* and *z*. If *y* and *z* are orthogonal (denoted $y \perp z$) then the angle between them is 90°. The cosine of 90° is zero, so $y \perp z \Rightarrow y'z = 0$. In terms of Figure 2, if the rays through A and J are orthogonal then the projection of J onto A (point C) is the origin.

[21] Technical efficiency can also be defined in terms of non-radial movements in inputs and outputs – see Fare, Grosskopf and Lovell (1985, ch.7). These alternative technical efficiency measures will not be discussed in this paper. In this paper, non-radial movements in inputs and outputs will be accommodated using measures of scale and mix efficiency.

[22] The Laspeyres and Paasche output aggregator functions are examples of linear aggregator functions.

outputs until it reaches point C. When written in terms of vector norms, the output-oriented measure of technical efficiency defined in equation (4.1) is $OTE_t = Q_t / \bar{Q}_t = \|A\| \|C\|$. If the linear aggregator functions are non-negative and non-decreasing in quantities then, irrespective of the values of the parameters of the aggregator functions, output- and input-oriented measures of efficiency will be given by the ratios of aggregate outputs in (4.1).

Figures 3 and 4 provide useful insights into technical efficiency measurement in the two-output two-input case. However, for firms that produce many outputs using many inputs, an alternative graphical representation is required. In this paper I have chosen to map feasible input and output combinations into aggregate quantity space. Figure 5 presents such a mapping for the input-output combinations represented by points A, B and C in Figures 3 and 4. In Figure 5, the point A represents the input-output combination (x_t, q_t), while the curve passing through points B and C represents the frontier of a *restricted* production possibilities set. The production possibilities set is restricted insofar as it only contains (aggregates of) input and output vectors that can be written as scalar multiples of x_t and q_t. Output-oriented technical efficiency is a measure of the vertical distance from point A to point C, while input-oriented technical efficiency is a measure of the horizontal distance from point A to point B.

Figure 5 is especially important because it illustrates that measures of technical efficiency can be written as ratios of measures of TFP, and therefore as ratios of tangent functions of angles. This idea underpins the TFP and profitability decompositions presented later in the paper and already mentioned in connection with Figure 1. Figure 5 is also important because it shows how the production choices available to a multiple-input multiple-output firm can be represented on a single graph. However, for reasons given later in this section, Figure 5 cannot always be used to define and graphically illustrate efficiency concepts in the same way similar figures are used in the case of single-input single-output firms.

Scale Efficiency

It is clear from Figure 5 that improvements in technical efficiency imply increases in TFP, and yet the TFP of Firm A is not maximized by moving to either of the technically efficient points B or C. If the input and output mixes are held fixed, Firm A will maximize its TFP by moving to a point where a ray through the origin is tangent to the restricted production possibilities frontier[23]. This point is represented by point D in Figure 6, and will be referred to as the point of *mix-invariant optimal scale* (MIOS). Pure scale efficiency is a measure of the difference between TFP at a technically efficient point and TFP at the point of MIOS. I use the term *pure* here because input and output mixes are being held fixed, so this change in TFP is a pure scale effect. Later I will define a scale efficiency measure that is contaminated by changes in the output and/or input mix. For now, measures of pure *output-oriented scale efficiency* (OSE) and pure *input-oriented scale efficiency* (ISE) are given by

(4.2) $$OSE_t = \frac{\bar{Q}_t / X_t}{\tilde{Q}_t / \tilde{X}_t} \quad \text{and} \quad ISE_t = \frac{Q_t / \bar{X}_t}{\tilde{Q}_t / \tilde{X}_t}$$

where $\tilde{Q}_t$ and $\tilde{X}_t$ denote the (output-mix and input-mix preserving) aggregate output and input quantities at the point of MIOS. Technically, $\tilde{Q}_t = \tilde{\lambda}_t Q_t$ and $\tilde{X}_t = \tilde{\rho}_t X_t$ where

(4.3) $$(\tilde{\lambda}_t, \tilde{\rho}_t) = \underset{\lambda>0,\rho>0}{\arg\max} \left\{ \lambda / \rho : (\rho x_t, \lambda q_t) \in T^t \right\}.$$

Two observations are in order. First, it is clear from Figure 6 that, like the measures of technical efficiency given by equation (4.1), the measures of scale efficiency given by equation (4.2) can be written as ratios of measures of TFP (slopes of rays through the origin). Again, this is fundamentally important for the TFP decompositions presented in

[23] Or, in the case of a constant returns to scale technology, a point where a ray through the origin coincides with the frontier.

Section 5. Second, the technology constraint in the optimization problem (4.3) can be written $\lambda D_O^t(\rho x_t, q_t) \leq 1$. This constraint will be binding at the optimum, which means the problem (4.3) reduces to the problem of finding the value ρ that minimizes $D_O^t(\rho x_t, \rho q_t)$. Thus, we can write:[24]

$$(4.4) \qquad OSE_t = \frac{D_O^t(\tilde{\rho} x_t, \tilde{\rho} q_t)}{D_O^t(x_t, q_t)} = \frac{\inf_\rho D_O^t(\rho x_t, \rho q_t)}{D_O^t(x_t, q_t)}$$

which corresponds to the measure of output scale efficiency defined by Balk (2001, p.164). See Balk (2001) for related discussion concerning the elasticity of scale.

Mix Efficiency

The efficiency measures discussed to this point have been defined with reference to a *restricted* production frontier – consideration has only been given to (aggregates of) input and output vectors that can be written as scalar multiples of x_t and q_t. Mix efficiency is a measure of the change in productivity when restrictions on the input and output mixes are relaxed. I use the term *mix* efficiency instead of *allocative* efficiency to avoid confusion with similar but potentially distinct value-based allocative efficiency measures defined later in the paper.

Relaxing restrictions on input and/or output mix leads to an expansion in the production possibilities set. The boundary of this expanded set is an *unrestricted* production frontier that envelops restricted frontiers of the type depicted in Figure 6. To illustrate the way in which the production possibilities set expands, again consider the special case where firm A uses $x_t = (x_{1t}, x_{2t})'$ to produce $q_t = (q_{1t}, q_{2t})'$ and where the input and output aggregator functions are linear. Recall that Figures 3 and 4 were previously used to depict measures of technical efficiency in this special case. Figures 7 and 8 are identical to Figures 3 and 4 except they have been embellished with representations of input mix and output mix effects. Recall that if both the input mix and output vector are held fixed then firm A can minimize aggregate input use by radially contracting inputs to point B in Figure 7. If restrictions on input mix are then relaxed, firm A can further reduce aggregate input use by moving to point U. This corresponds to a horizontal movement to a point that lies to the left of point B in Figure 6. Figure 8 depicts a similar expansion in output space. If the input vector and the output mix are held fixed, the best firm A can do is move to point C in Figure 8. However, if restrictions on output mix are removed, Firm A can further increase aggregate output by moving to point V in Figure 8, which corresponds to a vertical movement to somewhere above point C in Figure 6. Figure 9 is a version of Figure 6 that represents the unrestricted production frontier as the curved line passing through points U and V.

Pure mix efficiency is a measure of the difference between TFP at a technically efficient point on the (mix-) restricted frontier, and TFP at a point on the unrestricted frontier, holding either the input vector or the output vector fixed. Again, I use the term *pure* here because either the input vector or the output vector (i.e., the scale) is fixed, so this change in TFP is a pure mix effect. In terms of Figure 9, pure *output-oriented mix efficiency* (OME) is the difference between TFP at points C and V, while pure *input-oriented mix efficiency* (IME) is the difference between TFP at points B and U. Mathematically:

$$(4.4) \qquad OME_t = \frac{\bar{Q}_t}{\hat{Q}_t} \qquad \text{and} \qquad IME_t = \frac{\hat{X}_t}{\bar{X}_t}$$

where $\hat{Q}_t = Q(\hat{q}_t)$ and $\hat{X}_t = X(\hat{x}_t)$ are aggregates of

[24] Since the technology constraint is binding at the optimum, $\tilde{\lambda} = D_O^t(\tilde{\rho} x_t, q_t)^{-1}$. Equation (4.4) follows from the linear homogeneity of the output distance function and the definitions of $\bar{Q}_t, \tilde{Q}_t$ and $\tilde{X}_t$.

$$\hat{q}_t = \underset{q>0}{\arg\max}\left\{Q(q) : (x_t, q) \in T^t\right\} \quad \text{and} \quad \hat{x}_t = \underset{x>0}{\arg\min}\left\{X(x) : (x, q_t) \in T^t\right\}. \tag{4.5}$$

Figures 7, 8 and 9 can together be used to illustrate a fundamentally important property of an aggregate quantity representation of the production technology, namely that many different input-output combinations can map to the same point in aggregate quantity space. For example, *all* firms operating on *either* the dashed line through point A in Figure 7 *or* on the dashed line through point A in Figure 8 plausibly map to the single point A in Figure 9. Equally important is the fact that, except in restrictive special cases, *only* the firm operating at point A in Figure 3 *and* at point A in Figure 4 will face the restricted production possibilities frontier depicted by the curved line passing through points B and C in Figure 9. For example, a different firm operating at point R in Figure 7 is an input-mix efficient firm that will be represented by point A in Figure 9, but will face a restricted production possibilities frontier that *touches* the unrestricted frontier at point U. Likewise, an output-mix efficient firm operating at point S in Figure 8 will face a restricted production possibilities frontier that touches the unrestricted frontier in Figure 9 at point V. In more mathematical terms, the restricted frontiers associated with the points (x_t, q_t) and (x_s, q_s) may differ because the equalities $X(x_t) = X(x_s)$ and $Q(x_t) = Q(x_s)$ are not enough to ensure equality between the restricted production possibilities sets $\{(\lambda x_t, \rho q_t) : (\lambda x_t, \rho q_t) \in T^t, \lambda > 0, \rho > 0\}$ and $\{(\lambda x_s, \rho q_s) : (\lambda x_s, \rho q_s) \in T^t, \lambda > 0, \rho > 0\}$. Relatedly, even if the production technology exhibits constant returns to scale, the existence of scope economies (the mix effect). means that the unrestricted frontier in Figure 9 will not generally be a straight line.

Residual Scale Efficiency

Observe from Figure 9 that improvements in technical and mix efficiency imply increases in TFP, but the TFP of firm A is not maximized by moving to either of the technically- and mix-efficient points V or U. Rather, Firm A will maximize its TFP by moving to a point where a ray through the origin is tangent to the *unrestricted* production possibilities frontier. This point is represented by point E in Figure 10, and will be referred to as the point of *maximum productivity* (MP). Residual scale efficiency is a measure of the difference between TFP at a technically- and mix-efficient point and TFP at the point of MP. I use the term *scale* because any movement around an unrestricted production frontier is a movement from one mix-efficient point to another, so any improvement in TFP is essentially a scale effect. However, I also use the term *residual* because, even though all the points on the unrestricted frontier are mix-efficient, they may nevertheless have different input and output mixes (e.g., points V and U)[25]. Thus, what is essentially a measure of scale efficiency may contain a residual mix effect. The term *residual* is also appropriate in the sense that if we are interested in decomposing the difference between TFP at the observed point A and TFP at the point of maximum productivity E, then residual scale efficiency is the component that remains after we have accounted for pure technical and pure mix efficiency effects. Mathematically, measures of *residual output-oriented scale efficiency* (ROSE) and *residual input-oriented scale efficiency* (RISE) are

$$ROSE_t = \frac{\hat{Q}_t / X_t}{Q_t^* / X_t^*} \quad \text{and} \quad RISE_t = \frac{Q_t / \hat{X}_t}{Q_t^* / X_t^*} \tag{4.6}$$

where $Q_t^* = Q(q_t^*)$ and $X_t^* = X(x_t^*)$ are aggregates of

$$(x_t^*, q_t^*) = \underset{x>0, q>0}{\arg\max}\left\{Q(q) / X(x) : (x, q) \in T^t\right\} \tag{4.7}$$

That is, Q_t^* and X_t^* are the aggregate output and input quantities at the point of maximum productivity. TFP at that point is denoted $TFP_t^* = Q_t^* / X_t^*$.

[25] Observe from Figure 9 that the aggregate input at point V is the same as the aggregate input at point A, so Firm V could easily have the same input mix as Firms A and B. However, from Figure 7, the input mix at point B differs from the input mix at point U. .

Residual Mix Efficiency

Residual mix efficiency (RME) is a measure of the difference between TFP at the point of mix-invariant optimal scale (MIOS) and TFP at the point of maximum productivity (MP):

$$(4.8) \qquad RME_t = \frac{\tilde{Q}_t / \tilde{X}_t}{Q_t^* / X_t^*}$$

where the aggregate quantities in this equation have already been defined. This change in TFP is represented in Figure 11 by a movement from point D on the restricted (mix-invariant) production frontier to point E on the unrestricted frontier. The use of the term *mix* is self-evident – the movement from point D to point E is a movement from an optimal point on a mix-restricted frontier to an optimal point on an unrestricted frontier, so the difference in TFP is essentially a mix effect. Again, I use the term residual because i) the movement from point D to point E may also involve a change in scale, and ii) in the context of comparing TFP at points A and E, this measure can be viewed as the component that remains after accounting for pure technical and pure scale efficiency effects.

Other Efficiency Measures

Several other efficiency measures are defined in the productivity literature, but none are vital to the decomposition of either productivity or profitability change. Among these are measures of cost efficiency, revenue efficiency, profit efficiency, and associated measures of allocative efficiency. *Cost efficiency* (CE) is essentially the ratio of minimum cost to observed cost, *revenue efficiency* (RE) is the ratio of observed revenue to maximum revenue, and *profit efficiency* (PE) is the ratio of observed profit to maximum profit. More formally,

$$(4.9) \qquad CE_t = \frac{C^t(w_t, q_t)}{W_t X_t}$$

$$(4.10) \qquad RE_t = \frac{P_t Q_t}{R^t(p_t, x_t)}$$

$$(4.11) \qquad PE_t = \frac{P_t Q_t - W_t X_t}{\pi^t(p_t, w_t)}$$

where it is noteworthy that $P_t Q_t = p_t' q_t$ and $W_t X_t = w_t' x_t$ owing to the fact that the quantity-price aggregator function pairs satisfy the product rules. To define two associated measures of allocative efficiency, recall that $\bar{X}_t$ denotes the minimum aggregate input possible when using a scalar multiple of x_t to produce q_t, and $\bar{Q}_t$ denotes the maximum aggregate output that is possible when using x_t to produce a scalar multiple of q_t. Then *cost allocative efficiency* (CAE)[26] can be defined as the ratio of minimum cost to the cost of $\bar{X}_t$, while *revenue allocative efficiency* (RAE) can be defined as the inverse of the ratio of maximum revenue to the value of $\bar{Q}_t$. Mathematically:

$$(4.12) \qquad CAE_t = \frac{C^t(w_t, q_t)}{W_t \bar{X}_t} \qquad \text{and} \qquad RAE_t = \frac{P_t \bar{Q}_t}{R^t(p_t, x_t)}$$

A simple algebraic manipulation of equations (4.1), (4.9) and (4.10) reveals that the measures of cost and revenue efficiency given by (4.9) can be decomposed into the product of measures of technical efficiency and allocative efficiency:

[26] Farrell (1957) uses the term *price efficiency*.

(4.13) $\quad CE_t = ITE_t \times CAE_t \quad$ and $\quad RE_t = OTE_t \times RAE_t$

To help visualise these concepts, the cost efficiency decomposition is depicted in Figure 12. The points A, B and U in Figure 12 correspond to the points A, B and U in Figure 7, which was previously used to depict the concept of mix efficiency for a two-input firm where inputs were aggregated using the linear aggregator function $X(x_t) = \beta_1 x_{1t} + \beta_2 x_{2t}$. Recall that Firm A was able to increase its TFP by moving from point A to the technically-efficient but mix-inefficient point B, and was able to maximise its TFP by moving from point B to the technically- and mix-efficient point U. Each of these movements was associated with a reduction in aggregate input use, and Figure 12 now confirms that each of these movements is associated with a reduction in cost. In Figure 12, the dashed lines passing through points A, B and L are isocost lines with the same slope $-w_{1t}/w_{2t}$ but different intercepts: $W_t X_t / w_{2t} > W_t \bar{X}_t / w_{2t} > C^t(w_t, q_t)/w_{2t}$ which implies $W_t X_t > W_t \bar{X}_t > C^t(w_t, q_t)$. Figures 7 and 12 together reveal that i) Firm A can minimize cost by moving to the technically- and allocatively-efficient point L, ii) mix-efficient points of production do not necessarily coincide with allocatively-efficient points of production, and iii) a mix-efficient firm will also be allocatively-efficient, and vice versa, if the input aggregator function takes the Paasche form: $X(x_t) = w_{1t} x_{1t} + w_{2t} x_{2t}$. It is straightforward to map allocatively-efficient points such as point L to points in aggregate quantity space: point L in Figure 12 maps to a point on the horizontal line segment UB in Figure 9; on the revenue side, an allocatively-efficient point would map to a point on the vertical line segment VC in Figure 9.

One last visual example is provided in the form of Figure 13, which depicts the profit efficiency measure in aggregate output space. The points A, B, C, E, U and V in Figure 13 all correspond to the same points in Figures 9 to 11. In Figure 13, the dashed lines passing through points A and K are isoprofit lines with the same slope W_t / P_t but different intercepts. Among other things, Figure 13 illustrates that i) Firm A maximizes profit at the technically-efficient but scale-inefficient point K, ii) the profit-maximizing point of production does not necessarily coincide with the TFP-maximizing point of production, iii) the profit efficiency of Firm A can be decomposed into a component measuring the change in profits as Firm A moves to maximize TFP (at point E), and a further change in profits as Firm A moves to maximize profits (at point K), and iv) the profit- and TFP-maximizing points coincide (at point E) if and only if maximum TFP (the slope of the ray through point E) equals the inverse of the terms of trade (the slope of the isoprofit line) (i.e., if $Q_t^* / X_t^* = W_t / P_t = \tan e$).

5. DECOMPOSING PRODUCTIVITY AND PROFITABILITY CHANGE

Any input-output combination can be mapped to a point in aggregate quantity space. Such a mapping is useful because i) it provides for a single two-dimensional graphical representation of the production opportunities available to multiple-input multiple-output firms, and ii) multiplicatively complete indexes of total factor productivity and conventional ratio-type measures of efficiency can be expressed very simply in terms of angles. Both of these properties provide for any number of conceptually and mathematically simple decompositions of TFP change. For example, the previous section examined the case of a firm that selected the input-output combination (x_t, q_t) from the period-*t* production possibilities set T^t. That firm was mapped to point A in Figures 5, 6, 9, 10 and 11. In terms of aggregate quantities and angles, the TFP of this firm was measured as $TFP_t = Q_t / X_t = \tan a$; the maximum TFP possible holding the input vector and output mix fixed was $\bar{Q}_t / X_t = \tan c$; the maximum TFP possible holding the input vector fixed but allowing the output mix to vary was $\tilde{Q}_t / X_t = \tan v$; and the maximum TFP possible using *any* technically feasible inputs and outputs was $TFP_t^* = Q_t^* / X_t^* = \tan e$. These definitions imply that the difference between TFP at the point (x_t, q_t) and TFP at the point of maximum productivity can be decomposed as:

(5.1) $$TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{\tan a}{\tan e} = \frac{\tan a}{\tan c}\frac{\tan c}{\tan v}\frac{\tan v}{\tan e}. \qquad \text{(TFP efficiency)}$$

Of course, there are at least as many ways to decompose so-called *TFP efficie*ncy as there are points in the period-*t* production possibilities set. This particular decomposition is an output-oriented decomposition insofar as the input vector is, as far as possible, held fixed. In visual terms, it traces out the path A-C-V-E in Figure 13.

The intermediate points C and V were chosen for this example only because they feature in definitions of Farrell-type measures of technical and mix efficiency. It is clear that if the same type of decomposition is used for a firm that selects the input-output combination (x_0, q_0) from the period-0 production possibilities set, then the ratio-type productivity index $TFP_{0t} = TFP_t / TFP_0$ can be decomposed into output-oriented measures of technical and mix efficiency, plus a measure of the difference between points of maximum productivity. In this section I consider a small number of such decompositions and their relationships to other TFP decompositions in the literature.

Productivity Change

The decomposition given by equation (5.1) is one of two output-oriented decompositions of TFP efficiency that utilise the efficiency measures defined in Section 4. Equation (5.1) traces out the path A-C-V-E in Figure 13, while the second decomposition traces out the path A-C-D-E in Figures 11 and 13. In terms of the measures of efficiency defined in equations (4.1) to (4.8), these two output-oriented decompositions are:

(5.2) $$TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{Q_t / X_t}{Q_t^* / X_t^*} = OTE_t \times OME_t \times ROSE_t$$ and

(5.3) $$TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{Q_t / X_t}{Q_t^* / X_t^*} = OTE_t \times OSE_t \times RME_t$$

Geometric averages of equations (5.2) and (5.3) are also valid, as are geometric averages of their input-oriented counterparts. Importantly, analogous relationships hold for the reference firm 0. Even if we restrict our attention to this small number of decompositions, it is clear that the TFP index $TFP_{0t} = TFP_t / TFP_0$ can be decomposed in very many valid ways. Four possibilities are:

(5.4) $$TFP_{0t} = \left(\frac{TFP_t^*}{TFP_0^*}\right)\left(\frac{OTE_t}{OTE_0}\right)\left(\frac{OME_t}{OME_0}\right)\left(\frac{ROSE_t}{ROSE_0}\right)$$

(5.5) $$TFP_{0t} = \left(\frac{TFP_t^*}{TFP_0^*}\right)\left(\frac{OTE_t}{OTE_0}\right)\left(\frac{OSE_t}{OSE_0}\right)\left(\frac{RME_t}{RME_0}\right)$$

(5.6) $$TFP_{0t} = \left(\frac{TFP_t^*}{TFP_0^*}\right)\left(\frac{ITE_t}{ITE_0}\right)\left(\frac{IME_t}{IME_0}\right)\left(\frac{RISE_t}{RISE_0}\right)$$

(5.7) $$TFP_{0t} = \left(\frac{TFP_t^*}{TFP_0^*}\right)\left(\frac{ITE_t}{ITE_0}\right)\left(\frac{ISE_t}{ISE_0}\right)\left(\frac{RME_t}{RME_0}\right)$$

The output-oriented decompositions given by (5.4) and (5.5) are likely to be most relevant in empirical contexts where inputs are held fixed (i.e., contexts where researchers tend to estimate output-distance functions). The input-oriented decompositions (5.6) and (5.7) are likely to be most meaningful in industries where outputs are held fixed (i.e., contexts where researchers tend to estimate input-distance and cost functions). In contexts where inputs and outputs are both free to vary, geometric averages of any of the output- and input-oriented decompositions are available. Of course, if the technology exhibits constant returns to scale, then many of the input-oriented components are equal to the output-oriented components.

Common to the decompositions (5.4) to (5.7) is the ratio TFP_t^* / TFP_0^*. This term measures the difference between i) the maximum productivity possible when choosing from the production possibilities set T^t and ii) the maximum productivity possible when choosing from the production possibilities set T^0. Thus, it is a natural measure of technical change. In terms of Figure 11, it measures the change in the slope of the ray passing through point E. In contrast, in their decomposition of the Malmquist TFP index, Fare, et al. (1994) measure the change in the slope of the ray passing through point D. Thus, the Fare, et al. (1994) measure of technical change includes a mix effect and will typically vary from firm to firm.

The remaining terms in equations (5.4) to (5.7) are ratio measures of technical efficiency change, pure mix efficiency change, pure scale efficiency change, and residual measures of mix and scale efficiency change. Importantly, the decompositions given by these equations are *complete* in the sense that there is no unexplained component. Equally importantly, these equations were derived without any assumptions concerning firm optimising behaviour, the structure of markets, or returns to scale. They can be used in cases where prices are determined by market competition, by the firm, or by a regulator, and they can be used when technologies exhibit constant or variable returns to scale. In theory, any multiplicatively complete TFP index can be decomposed in this way (and any additively complete TFP index can be decomposed in a similar way)

Profitability Change

Recall from Section 3 that profitability change can be decomposed into the product of a TFP index and an index measuring the change in the terms of trade (see equation 3.16). The decompositions given by equations such as (5.4) to (5.7) now provide for any number of meaningful decompositions of profitability change. For example, multiplying the terms of trade index by the geometric average of equations (5.4) and (5.6) leads to the following decomposition:

$$(5.10)\qquad \Pi_{0t} = \left(\frac{P_{0t}}{W_{0t}}\right)\left(\frac{TFP_t^*}{TFP_0^*}\right)\left(\frac{OTE_t}{OTE_0}\frac{ITE_t}{ITE_0}\right)^{1/2}\left(\frac{OME_t}{OME_0}\frac{IME_t}{IME_0}\right)^{1/2}\left(\frac{ROSE_t}{ROSE_0}\frac{RISE_t}{RISE_0}\right)^{1/2}$$

Decompositions of profit change are also available. To get a sense of the difference between profit and profitability change, let the price and quantity vectors (p_0, w_0) and (x_0, q_0) be mapped to aggregate price and quantity points G and H superimposed on Figure 2. Profitability change would then be measured as $\Pi_{0t} = (\tan a/\tan g)(\tan j/\tan f)$, whereas profit change would be measured as $\pi_t - \pi_0 = \|A\|\|J\|\cos(180 - a - j) - \|G\|\|F\|\cos(180 - h - f)$.

Growth Accounting

The term growth accounting is usually associated with the work of Solow (1957) and is mainly used in economics to refer to models where measures of economic growth are decomposed into measures of input growth and technical change. Any decomposition of this type necessitates a description of the production technology, and a logical starting point is an aggregate production function of the form

$$(5.11)\qquad Q_t \le f(x_t, t)$$

where the inequality sign reflects possible inefficiency and $f(.)$ is, among other things, non-decreasing in inputs. The pioneering work of Solow (1957) was underpinned by the assumptions that technical change is Hicks-neutral and the technology exhibits constant returns to scale. Technical change is defined to be Hicks-neutral if and only if $f(x_t, t)$ can be written in the form $f(x_t, t) = A(t)X(x_t)$. The technology will exhibit constant returns to scale if the function $X(.)$ is linearly homogeneous. With these two assumptions equation (5.11) becomes

$$(5.12)\qquad Q_t \le A(t)X(x_t)$$

where $X(x_t)$ is a non-decreasing linearly homogenous (aggregator) function. It is clear that $A(t)$ only provides an upper bound on $TFP_t \equiv Q_t / X(x_t)$, so nothing can be said about TFP growth without being more explicit about the nature of inefficiency. The usual way forward is to implicitly assume away any technical and mix inefficiency. There is no scale inefficiency because the technology is already assumed to exhibit constant returns to scale. With these assumptions, equation (5.12) holds with equality and TFP growth can be measured using

$$(5.13) \qquad TFP_{0t} = \frac{A(t)}{A(0)}$$

This measure of TFP growth is frequently reported in the literature. Applied researchers should note carefully that (5.13) is only a legitimate measure of productivity growth if i) technical change is Hicks neutral, ii) the technology exhibits constant returns to scale, iii) firms are technically efficient, iv) firms are input-mix efficient, and v) firms are output-mix efficient. In any empirical application, the joint probability of meeting all of these requirements may be near zero. Of course, if these requirements are not met then TFP growth can be measured as the ratio of an output quantity index to an input quantity index, and decomposed using (geometric averages of) equations (5.4) to (5.7).

6. EMPIRICAL EXAMPLE

Coelli, Rao, O'Donnell and Battese (2005) use fifteen observations on hypothetical single-input single-output firms to illustrate the Fare, et al. (1994) decomposition of the Malmquist TFP index. This section uses similar data to illustrate the decomposition of the multiplicatively-complete Moorsteen-Bjurek index. The data is presented in Table 1 and comprises observations on the outputs and inputs of five hypothetical firms over four time periods.

Estimates of output- and input-oriented measures of technical, scale and mix efficiency are reported in Table 2. These estimates were computed using data envelopment analysis (DEA) programs that allow for variable returns to scale and technical regress[27]. They confirm patterns of efficiency that are also clearly evident in the data. It will be useful to consider these patterns one period at a time.

First, the period-1 results indicate that every firm is fully-mix efficient, reflecting the fact that all firms in this period have the same output mix and the same input mix (each firm produces half as much output 1 as output 2 and uses half as much input 1 as input 2). Firm 3 is also the only firm that is found to be fully technically-, mix- and scale-efficient, reflecting the fact that it produces more of any output per unit of any input than any other firm (it is the TFP-maximizing firm in this period). The plausibility of these efficiency estimates is also evident from Figure 14 where all period-1 observations are depicted in relation to the estimated variable returns to scale production frontier. This two-dimensional graphical representation is made possible by the fact that there are no mix effects, so that any aggregate output (input) will be proportional to the amount of output 1 (input 1)[28]. Because the amounts of output 1 and input 1 are identical to the period-1 outputs and inputs listed in Coelli, et al. (2005, p.296), Figure 14 is isomorphic to their Figure 11.1. Observe from Figure 14 that firms 1 and 5 are fully technically efficient, so, in view of the fact that they are also fully mix-efficient, the entire TFP shortfall is attributed to scale inefficiency. At the same time, the outputs and inputs of firms 2 and 4 place them inside the boundary of the production frontier, so TFP shortfalls are plausibly attributed to both technical inefficiency and scale inefficiency.

Second, observe that in period 2 all firms have been able to double their outputs without any increases in input use. It follows that the period-2 production frontier is isomorphic to the period-1 frontier depicted in Figure 14, and all measures of technical, scale and mix efficiency in the two periods must be identical.

Third, the inputs used by all firms in period 3 are the amounts that had been used by firm 3 in period 2 (the most productive firm in that period). However, different firms can be seen to produce outputs in different amounts

[27] See O'Donnell (2009) for the motivation and a full description of these and related DEA programs.

[28] The labels along the axes reflect the fact that aggregate inputs and outputs are only identified up to an arbitrary positive constant.

and proportions. Firm 3 is still the best-practice benchmark and is still found to be fully technically-, scale- and mix-efficient. Firm 1 is the only other firm that is found to be fully-mix efficient, due to the fact that it is the only other firm that has the same (TFP-maximizing) output mix as firm 3. However, firm 1 only produces half as much output as firm 3, so it is plausibly only 50% technically efficient All other firms have a different output mix to firm 3 and are therefore found to be mix-inefficient using an output-oriented measure. Observe that Firms 4 and 5 have the same output mix and are therefore found to be equally mix-inefficient.

Finally, in period 4, all firms are found to be fully output-mix and input-mix efficient, reflecting the fact that all firms produce the same outputs and have the same input mix (the two inputs are used in equal amounts). Firm 1 is found to be fully technically-, scale- and mix-efficient, reflecting the fact that it has now surpassed firm 3 as the most productive firm in the sample. The plausibility of the remaining technical and scale efficiency estimates can be established with the aid of Figure 15, where all observations are seen lie on the boundary of a rectangular variable returns to scale production frontier. From an output-oriented perspective, all firms are technically-efficient, so all shortfalls in TFP are attributed to scale inefficiency. From an input-oriented perspective, all firms are scale-efficient, so all shortfalls in TFP are attributed to excess input usage (i.e., technical inefficiency).

The efficiency estimates reported in Table 2 have been used to compute corresponding estimates of efficiency change, and these are reported in Table 3. Also reported in Table 3 are estimates of technical change, the Moorsteen-Bjurek index of TFP change, and estimates of residual efficiency change. The measure of technical change and the TFP index were computed using DEA problems presented in O'Donnell (2009), while the residual efficiency change measures were computed as precisely that – residuals. The values reported for each firm in each period are non-cumulative firm-specific index numbers, which is to say they measure period-on-period changes for individual firms. For example, the row corresponding to observation 20 reveals that the TFP of firm 5 in period 4 was only 42% of the TFP of that same firm in the previous period, and that this productivity decline was due to a large fall in efficiency.

Tables 4 and 5 present estimates of efficiency, efficiency change and TFP change under the assumption the technology exhibits constant returns to scale. The plausibility of these results is easily established with the aid of Figures 14 and 15, where the constant returns to scale frontiers are represented by the dashed rays passing through the origin and the points of maximum productivity. Observe from Table 4 that all firms are fully scale-efficient. By chance, estimates of mix efficiency are identical to the estimates obtained under the variable returns to scale assumption.

The technical change estimates reported in Tables 3 and 5 are noteworthy in several respects. First, observe that estimates of the rate of technical change are invariant to the returns to scale assumption. This is because the constant- and variable-returns-to-scale DEA problems identify a TFP-maximizing data point that is common to both frontiers. Second, it can be seen that, in any period, the measure of technical change is identical for every firm. A theoretical explanation for this outcome was provided in Section 5. It is an intuitively plausible outcome insofar as, in any period, every firm has access to the same production possibilities set, so expansions or contractions in that set should impact on all firms equally. Third, the technical change estimates reported in the first five rows of column B reveal that the maximum TFP possible in period 2 was twice the maximum TFP possible in period 1. This is consistent with our earlier observation that there were no changes in technical, scale or mix efficiency between these two periods, and yet firms were able to double their outputs without any increase in input use. Fourth, the estimates reported in the next five rows reveal there was no technical change between periods 2 and 3. This is consistent with our observation that firm 3 was the best-practice benchmark in both periods, and in each period it used the same inputs to produce the same outputs. Finally, the rate of technical progress between periods 3 and 4 was 12.5%. This is an estimate of technical change that accounts for the change in TFP-maximising input mix between the two periods.

For purposes of comparison, estimated Malmquist TFP index numbers and associated estimates of the components of TFP change are reported in Table 6. These estimates were obtained using the DEA methodology described in Fare, et al. (1994). The Fare, et al. (1994) methodology involves estimating the distance functions in equations (3.14) and (3.15) under the assumption that the technology exhibits constant returns to scale.

It is not surprising that the Malmquist index numbers reported in Table 6 differ from the Moorsteen-Bjurek index numbers reported in Table 5, even though both were computed under the assumption of constant returns to scale. After all, the Moorsteen-Bjurek index is multiplicatively complete, but the Malmquist index is not. Nor is it surprising that the estimated components of TFP change differ. The Moorsteen-Bjurek components are computed under the assumption that the technology exhibits constant returns to scale, but, even though the Malmquist TFP index is computed under that same assumption, the Malmquist measures of technical and scale efficiency change are computed under a variable returns to scale assumption. In fact, the Malmquist measures of technical and scale efficiency change are those used in the decomposition of the Moorsteen-Bjurek index when the technology is assumed to exhibit variable returns to scale. This partly explains why the Malmquist index numbers reported in Table 6 are identical to the Moorsteen-Bjurek index numbers reported in Table 3. Those two sets of numbers are identical because the technical change and scale efficiency change components are identical, and because combined measures of technical and mix efficiency change are by chance identical for this small data set. In general, the Malmquist TFP index will differ from the Moorsteen-Bjurek TFP index. For example, the equality we observe in this example can be broken by simply augmenting the data set with an additional input that takes the value one for every firm in every time period. Another example involving a large agricultural data set is found in O'Donnell (2009).

7. CONCLUSION

This paper uses an aggregate price-quantity framework to show that there exist large classes of TFP index numbers that can be decomposed into measures of technical change, technical efficiency change, mix efficiency change and scale efficiency change. Alternative decompositions involving measures of revenue, cost and profit efficiency change are also available, but each of these alternative decompositions involves at least one measure of efficiency that does not yet feature prominently in the economics literature. There are, in fact, at least as many decompositions of TFP growth as there are points in the production possibilities set, but only a handful of these exclusively contain components that are economically interesting. Importantly, none of these decompositions rely on assumptions concerning market structure, so they can be used in industries where prices are determined by market competition, regulatory agencies, or by the firm itself. Nor is there any requirement that the technology exhibit constant returns to scale.

An important contribution of the paper has been to explain why the Malmquist TFP index of Caves, et al. (1982a) may be an unreliable measure of TFP change. This is because it is neither additively nor multiplicatively complete. It is ironic that the "[Malmquist] index has achieved much greater popularity than the [Moorsteen-Bjurek] index [partly because] it decomposes into various sources of productivity change" (Lovell, 2003, p.438) and yet, unless the technology is inversely homothetic and exhibits constant returns to scale, it is the latter index, not the former, that can be decomposed in an economically-meaningful way. Perhaps this is the most important contribution of the paper – to show that, irrespective of the returns to scale and scope properties of the production technology, the Moorsteen-Bjurek index can be meaningfully decomposed into any number of measures of technical change and efficiency change. The paper demonstrates that several input- and output-oriented decompositions can be empirically implemented using the DEA programs in O'Donnell (2009).

Appendix A

Consider the following decomposition of $C_t - C_0$:

(A.1)
$$
\begin{aligned}
C_t - C_0 &= \sum_{n=1}^{N}\left(w_{nt}x_{nt} - w_{nt}x_{n0} + w_{nt}x_{n0} - w_{n0}x_{n0}\right) \\
&= \sum_{n=1}^{N}\left(w_{nt}x_{nt} - w_{nt}x_{n0} + w_{nt}x_{n0} - w_{n0}x_{n0}\right) \\
&= \sum_{n=1}^{N} w_{nt}\left(x_{nt} - x_{n0}\right) + \sum_{n=1}^{N}\left(w_{nt} - w_{n0}\right)x_{n0} \\
&= \sum_{n=1}^{N} w_{nt}x_{nt}\left(\frac{x_{nt} - x_{n0}}{x_{nt}}\right) + \sum_{n=1}^{N}\left(\frac{w_{nt} - w_{n0}}{w_{n0}}\right)w_{n0}x_{n0} \\
&= C_t\sum_{n=1}^{N} s_{nt}\left(\frac{x_{nt} - x_{n0}}{x_{nt}}\right) + C_0\sum_{n=1}^{N} s_{n0}\left(\frac{w_{nt} - w_{n0}}{w_{n0}}\right)
\end{aligned}
$$

Further manipulation yields:

(A.2)
$$
C_{0t} = \frac{C_t}{C_0} = \frac{\sum_{n=1}^{N} s_{n0}\left(\frac{w_{nt}}{w_{n0}}\right)}{\sum_{n=1}^{N} s_{nt}\left(\frac{x_{n0}}{x_{nt}}\right)} = \frac{\sum_{n=1}^{N} w_{nt}x_{n0} \Big/ \sum_{n=1}^{N} w_{n0}x_{n0}}{\sum_{n=1}^{N} w_{nt}x_{n0} \Big/ \sum_{n=1}^{N} w_{nt}x_{nt}} = W_{0t}^{L} X_{0t}^{P}
$$

The alternative decomposition

(A.3)
$$
C_t - C_0 = \sum_{n=1}^{N}\left(w_{nt}x_{nt} - w_{n0}x_{nt} + w_{n0}x_{nt} - w_{n0}x_{n0}\right)
$$

eventually leads to

(A.4)
$$
C_{0t} = W_{0t}^{P} X_{0t}^{L}
$$

Appendix B

Let prices and quantities be non-negative continuous functions of time so that, for example, cost is also a non-negative continuous function of time. To emphasise this continuous dependence on time I use the notation

$$\text{(B.1)} \qquad C(t) = \sum_{n=1}^{N} w_n(t) x_n(t)$$

Total differentiation of (B.1) with respect to time yields (e.g. Caves, Christensen and Swanson, 1980, p.168):

$$\text{(B.2)} \qquad \frac{d \ln C(t)}{dt} = \sum_{n=1}^{N} s_n(t) \frac{d \ln x_n(t)}{dt} + \sum_{n=1}^{N} s_n(t) \frac{d \ln w_n(t)}{dt}$$

where $s_n(t) \equiv w_n(t) x_n(t) / C(t)$. Equivalently, using Newton's notation[29]:

$$\text{(B.3)} \qquad \frac{\dot{C}}{C(t)} = \sum_{n=1}^{N} s_n(t) \frac{\dot{x}_n}{x_n(t)} + \sum_{n=1}^{N} s_n(t) \frac{\dot{w}_n}{w_n(t)}$$

It follows that

$$\text{(B.4)} \qquad \int_0^t \frac{\dot{C}}{C(u)} du = \int_0^t \left(\sum_{n=1}^{N} s_n(u) \frac{\dot{x}_n}{x_n(u)} \right) du + \int_0^t \left(\sum_{n=1}^{N} s_n(u) \frac{\dot{w}_n}{w_n(u)} \right) du = \ln \frac{X(t)}{X(0)} + \ln \frac{W(t)}{W(0)}$$

where $X(t)$ and $W(t)$ are the Divisia aggregate output quantities and prices in period *t* (e.g., Usher, 1974, p. 274, eqs. 1,2; Hulten, 1973, p.1017, footnote.2):

$$\text{(B.5)} \qquad X(t) = X(0) \exp \left\{ \int_0^t \left(\sum_{n=1}^{N} s_n(u) \frac{\dot{x}_n}{x_n(u)} \right) du \right\} \qquad \text{(Divisia aggregate input quantity)}$$

$$\text{(B.6)} \qquad W(t) = W(0) \exp \left\{ \int_0^t \left(\sum_{n=1}^{N} s_n(u) \frac{\dot{w}_n}{w_n(u)} \right) du \right\} \qquad \text{(Divisia aggregate input price)}$$

and where $X(0)$ and $W(0)$ are arbitrary (but commonly normalised to unity). Of course, for any function $C(t)$

$$\text{(B.7)} \qquad \int_0^t \frac{\dot{C}}{C(u)} du = \int_0^t \frac{1}{C(u)} \frac{dC(u)}{du} du = \int_0^t \frac{1}{C(u)} dC(u) = \left[\ln C(u) \right]_0^t = \ln \frac{C(t)}{C(0)}$$

Together, equations (B.4) and (B.7) imply that i) a simple cost index can be decomposed into the product of a Divisia input quantity index and a Divisia input price index, and ii) Divisia quantity and price indexes satisfy the factor reversal test.

[29] Newton's notation for differentiation, also called the dot notation, places a dot over the function name to represent a derivative. If $y = f(t)$ then $\dot{y}$ and $\ddot{y}$ denote the first and second derivatives of *y* with respect to *t*. This notation is used almost exclusively for time derivatives.

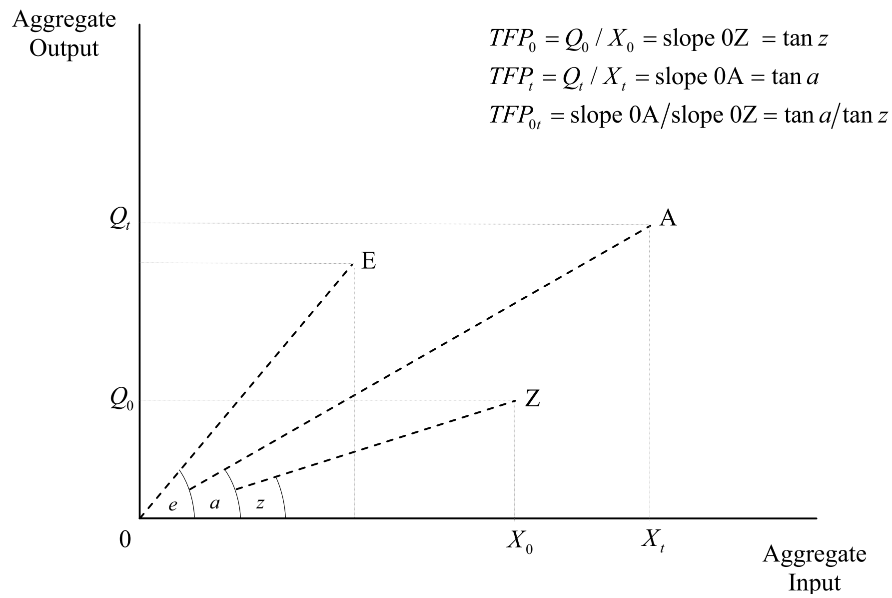

Figure 1. Total Factor Productivity Change

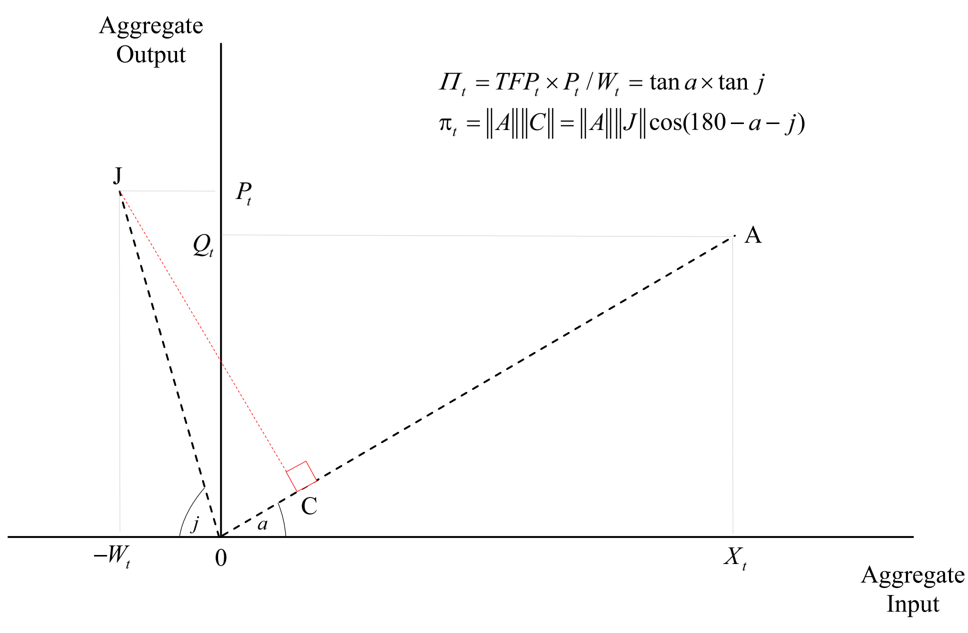

Figure 2. Profit and Profitability

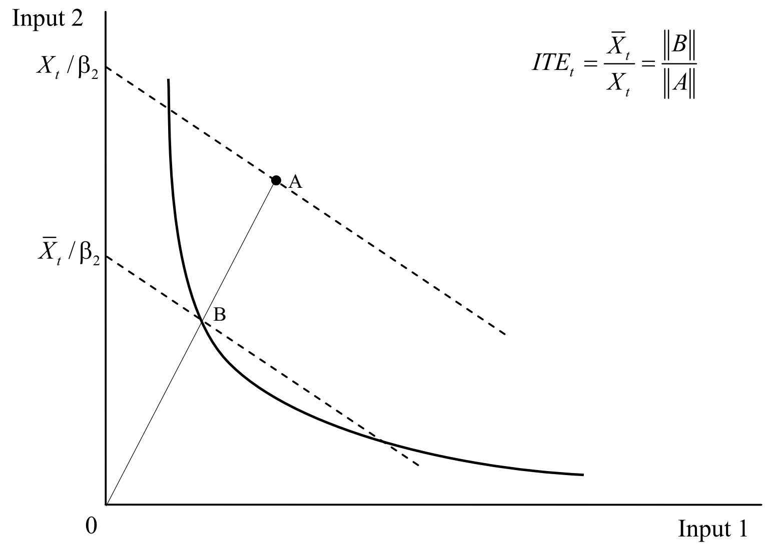

Figure 3. Input-Oriented Technical Efficiency for a Two-Input Firm

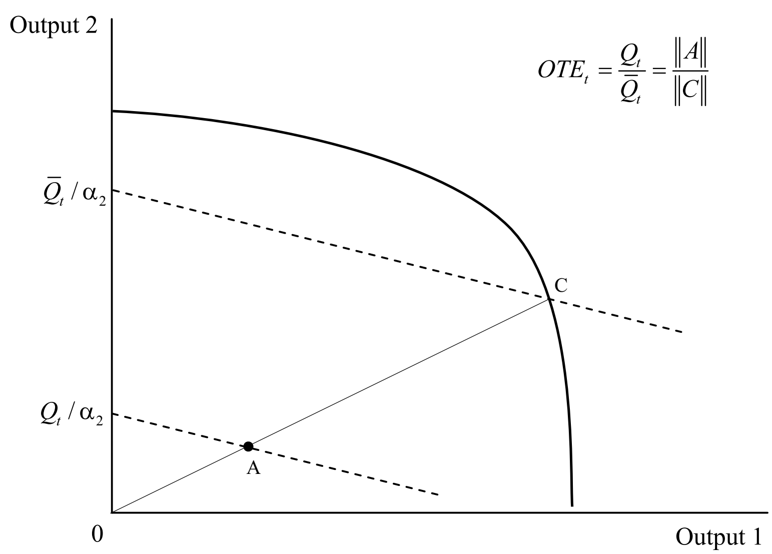

Figure 4. Output-Oriented Technical Efficiency for a Two-Output Firm

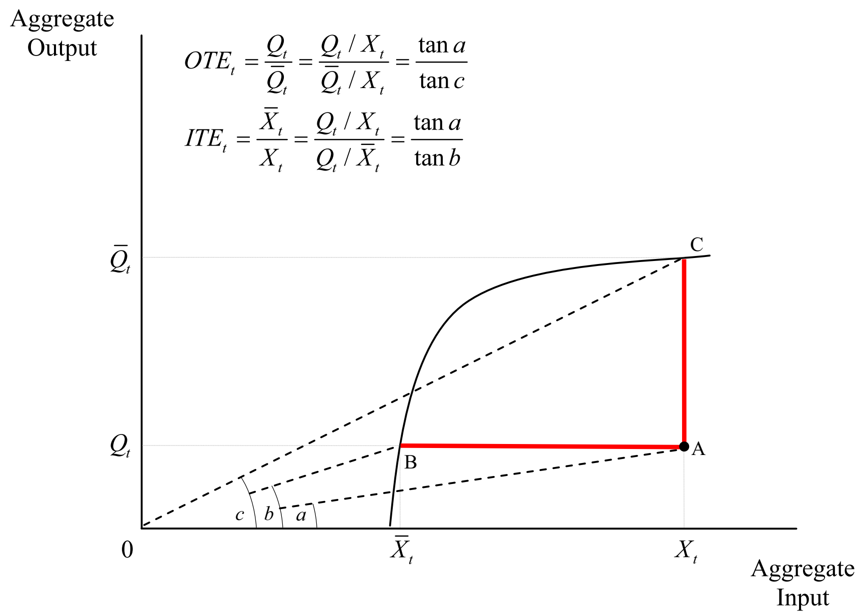

Figure 5. Input- and Output-Oriented Technical Efficiency for a Multiple-Input Multiple-Output Firm

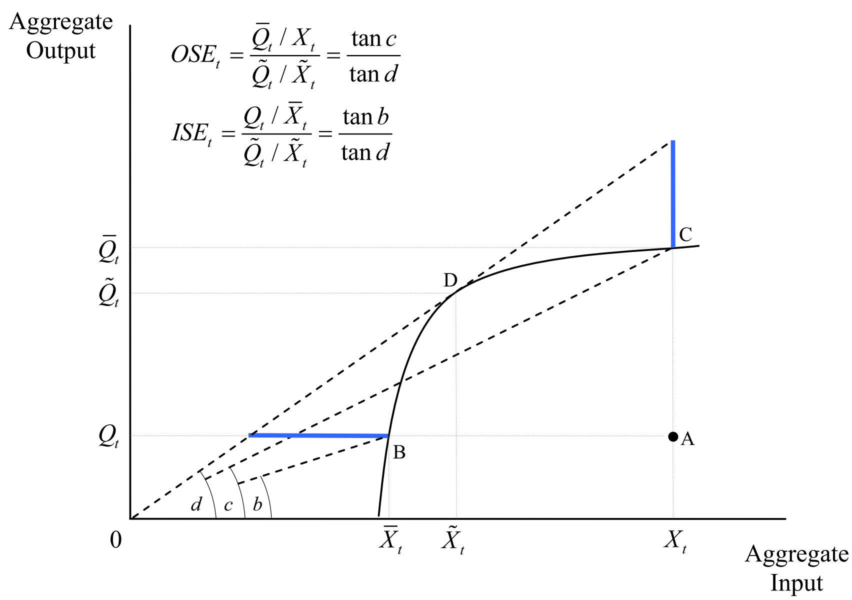

Figure 6. Output- and Input-Oriented Scale Efficiency for a Multiple-Input Multiple-Output Firm

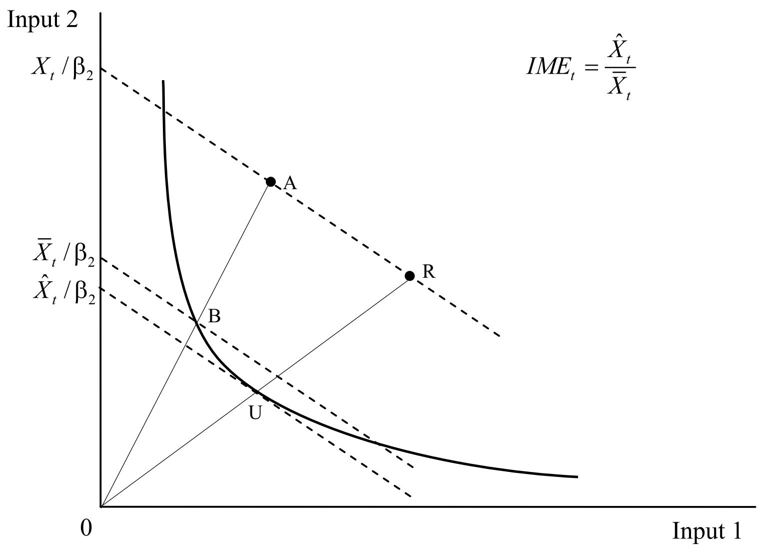

Figure 7. Input-Oriented Mix Efficiency for a Two-Input Firm

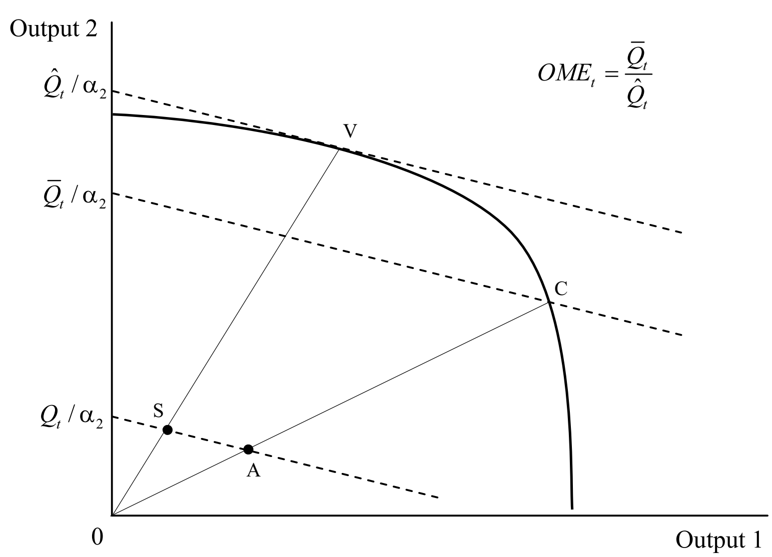

Figure 8. Output-Oriented Mix Efficiency for a Two Output Firm

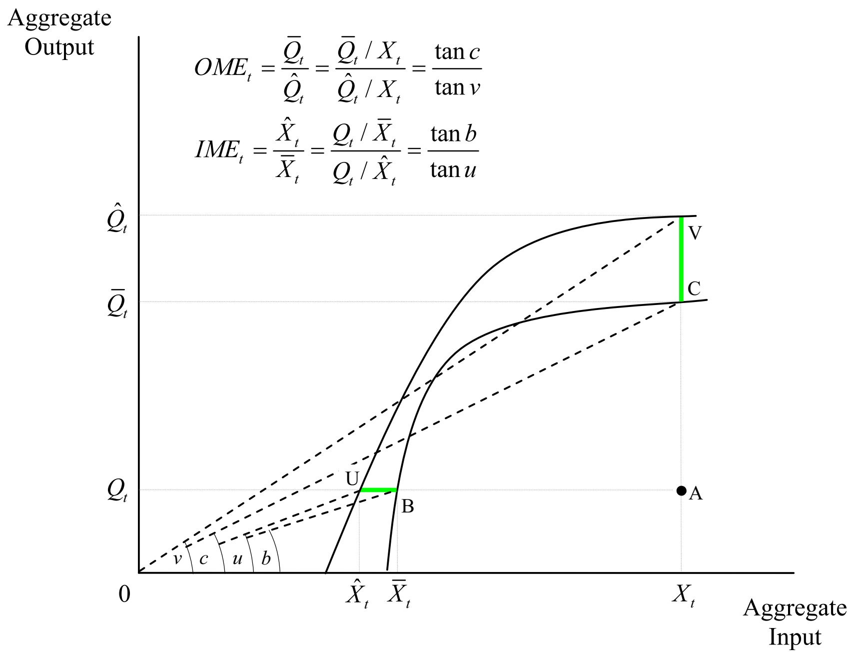

Figure 9. Output- and Input-Oriented Mix Efficiency for a Multiple-Input Multiple-Output Firm

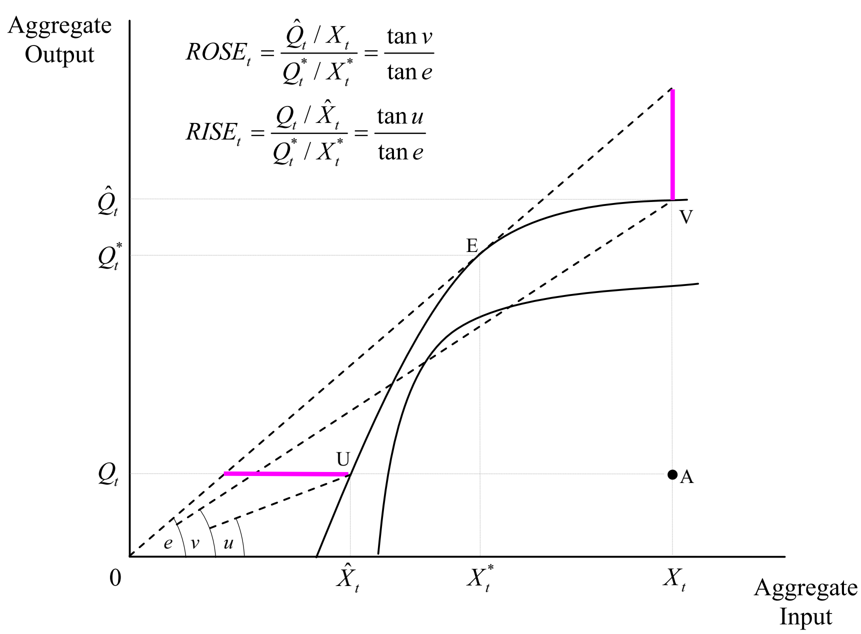

Figure 10. Residual Output- and Input-Oriented Scale Efficiency for a Multiple-Input Multiple-Output Firm

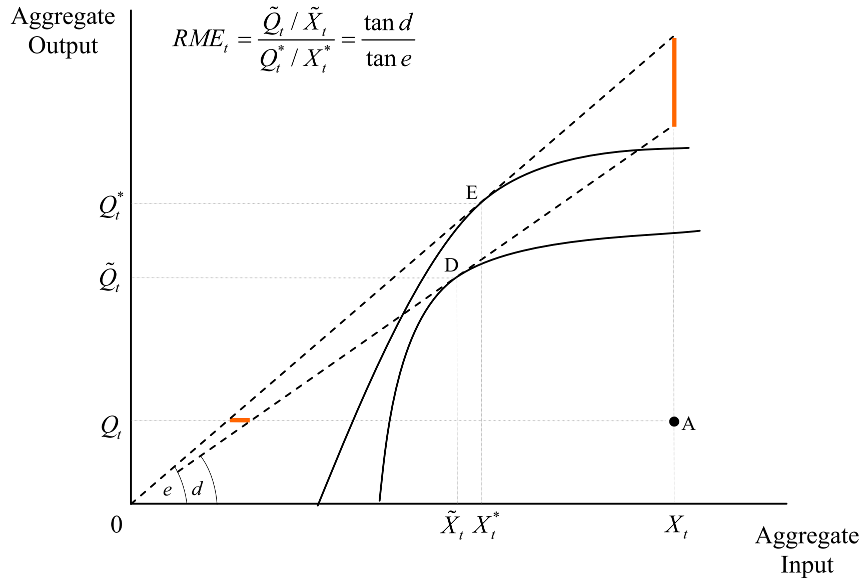

Figure 11. Residual Mix Efficiency for a Multiple-Input Multiple-Output Firm

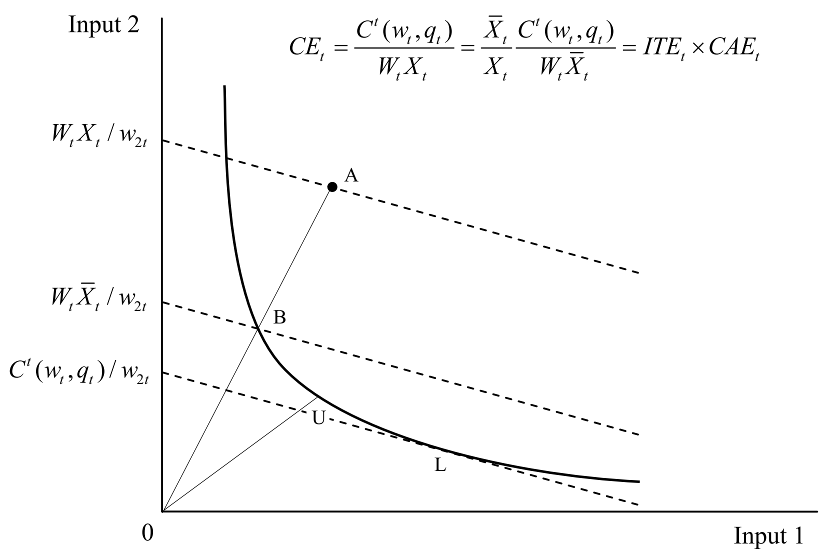

Figure 12. Cost and Cost-Allocative Efficiency for a Two-Input Firm

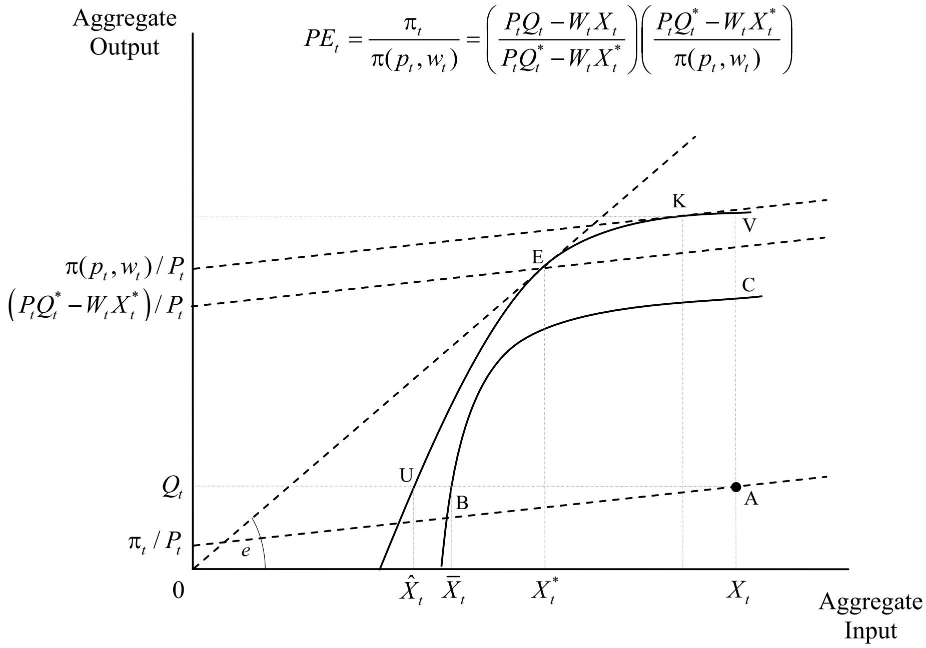

Figure 13. Profit Efficiency for a Multiple-Input Multiple-Output Firm

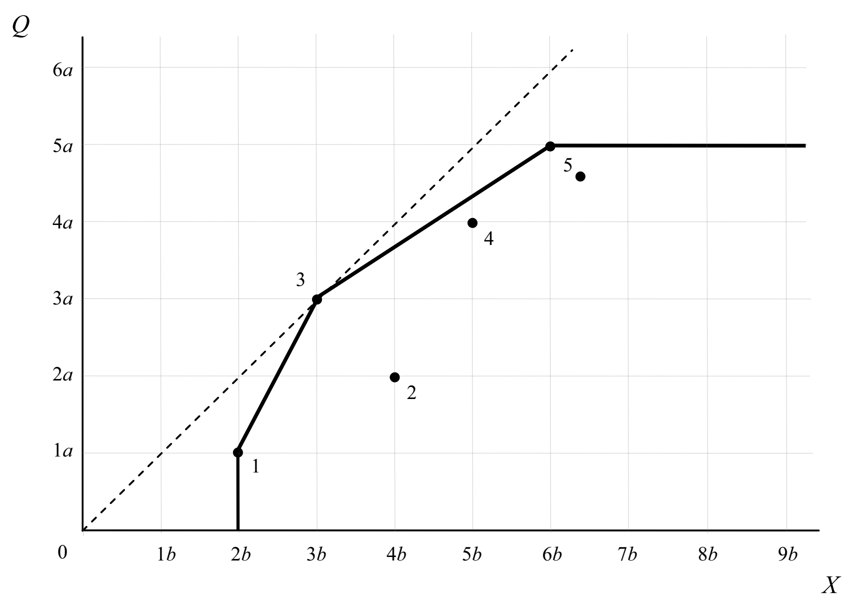

Figure 14. Observed Outputs and Inputs in Period 1

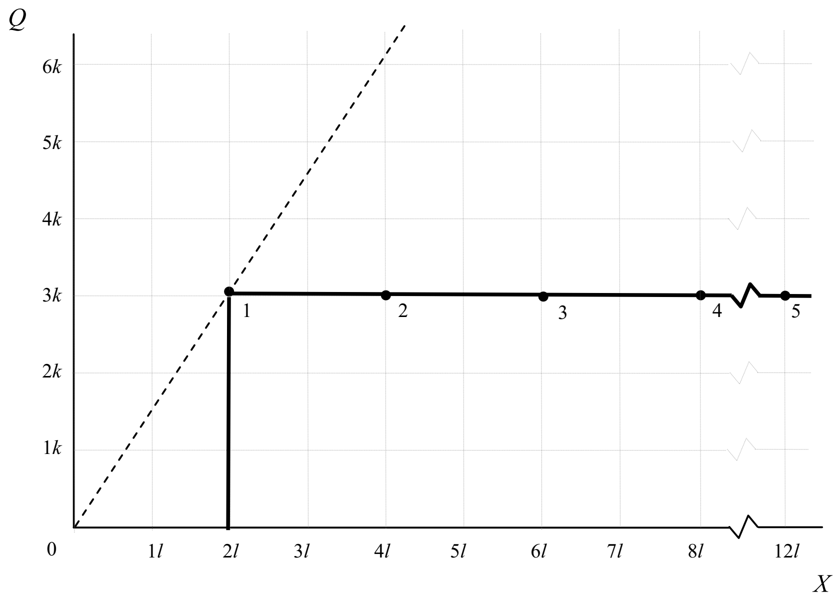

Figure 15. Observed Outputs and Inputs in Period 4

Table 1. DATA

Obs	Year	Firm	Outputs 1	Outputs 2	Inputs 1	Inputs 2
1	1	1	1	2	2	4
2	1	2	2	4	4	8
3	1	3	3	6	3	6
4	1	4	4	8	5	10
5	1	5	5	10	6	12
6	2	1	2	4	2	4
7	2	2	4	8	4	8
8	2	3	6	12	3	6
9	2	4	8	16	5	10
10	2	5	10	20	6	12
11	3	1	3	6	3	6
12	3	2	6	3	3	6
13	3	3	6	12	3	6
14	3	4	6	6	3	6
15	3	5	3	3	3	6
16	4	1	3	6	2	2
17	4	2	3	6	4	4
18	4	3	3	6	6	6
19	4	4	3	6	8	8
20	4	5	3	6	10	10

Table 2. MEASURES OF TECHNICAL, SCALE and MIX EFFICIENCY (ASSUMING VRS)

Obs	Year	Firm	OTE (D)	OSE (E)	OME (F)	ITE (H)	ISE (I)	IME (J)
1	1	1	1.000	0.5000	1.000	1.000	0.5000	1.000
2	1	2	0.5455	0.9167	1.000	0.6250	0.8000	1.000
3	1	3	1.000	1.000	1.000	1.000	1.000	1.000
4	1	4	0.9231	0.8667	1.000	0.9000	0.8889	1.000
5	1	5	1.000	0.8333	1.000	1.000	0.8333	1.000
6	2	1	1.000	0.5000	1.000	1.000	0.5000	1.000
7	2	2	0.5455	0.9167	1.000	0.6250	0.8000	1.000
8	2	3	1.000	1.000	1.000	1.000	1.000	1.000
9	2	4	0.9231	0.8667	1.000	0.9000	0.8889	1.000
10	2	5	1.000	0.8333	1.000	1.000	0.8333	1.000
11	3	1	0.5000	1.000	1.000	1.000	0.5000	1.000
12	3	2	1.000	1.000	0.5000	1.000	1.000	1.000
13	3	3	1.000	1.000	1.000	1.000	1.000	1.000
14	3	4	1.000	1.000	0.6667	1.000	1.000	1.000
15	3	5	0.5000	1.000	0.6667	1.000	0.5000	1.000
16	4	1	1.000	1.000	1.000	1.000	1.000	1.000
17	4	2	1.000	0.5000	1.000	0.5000	1.000	1.000
18	4	3	1.000	0.3333	1.000	0.3333	1.000	1.000
19	4	4	1.000	0.2500	1.000	0.2500	1.000	1.000
20	4	5	1.000	0.2000	1.000	0.2000	1.000	1.000
Firm	1	A	0.8409	0.7071	1.000	1.000	0.5946	1.000
Firm	2	B	0.7385	0.8051	0.8409	0.6648	0.8944	1.000
Firm	3	C	1.000	0.7598	1.000	0.7598	1.000	1.000
Firm	4	D	0.9608	0.6583	0.9036	0.6708	0.9428	1.000
Firm	5	E	0.8409	0.6105	0.9036	0.6687	0.7676	1.000
Year	1		0.8718	0.8016	1.000	0.8913	0.7841	1.000
Year	2		0.8718	0.8016	1.000	0.8913	0.7841	1.000
Year	3		0.7579	1.000	0.7402	1.000	0.7579	1.000
Year	4		1.000	0.3839	1.000	0.3839	1.000	1.000
Mean			0.8712	0.7047	0.9276	0.7431	0.8262	1.000
Minim			0.5000	0.2000	0.5000	0.2000	0.5000	1.000
Maxim			1.000	1.000	1.000	1.000	1.000	1.000

Table 3. DECOMPOSITION OF MOORSTEEN-BJUREK TFP INDEX (ASSUMING VRS)

						Components of Efficiency Change								
Obs	Year	Firm	TFP Index = (A)	Tech Change x (B)	Eff Change (C)	dOTE (D)	dOSE (E)	dOME (F)	dROSE (G)	dITE (H)	dISE (I)	dIME (J)	dRISE (K)	dRME (L)
6	2	1	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	2	2	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	2	3	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	2	4	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	2	5	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	3	1	1.000	1.000	1.000	0.5000	2.000	1.000	2.000	1.000	1.000	1.000	1.000	1.000
12	3	2	2.000	1.000	2.000	1.833	1.091	0.5000	2.182	1.600	1.250	1.000	1.250	1.000
13	3	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	3	4	1.250	1.000	1.250	1.083	1.154	0.6667	1.731	1.111	1.125	1.000	1.125	1.000
15	3	5	0.6000	1.000	0.6000	0.5000	1.200	0.6667	1.800	1.000	0.6000	1.000	0.6000	1.000
16	4	1	2.121	1.125	1.886	2.000	1.000	1.000	0.9428	1.000	2.000	1.000	1.886	0.9428
17	4	2	0.5303	1.125	0.4714	1.000	0.5000	2.000	0.2357	0.5000	1.000	1.000	0.9428	0.9428
18	4	3	0.3536	1.125	0.3143	1.000	0.3333	1.000	0.3143	0.3333	1.000	1.000	0.9428	0.9428
19	4	4	0.2652	1.125	0.2357	1.000	0.2500	1.500	0.1571	0.2500	1.000	1.000	0.9428	0.9428
20	4	5	0.4243	1.125	0.3771	2.000	0.2000	1.500	0.1257	0.2000	2.000	1.000	1.886	0.9428
Firm	1	A	1.619	1.310	1.235	1.000	1.260	1.000	1.235	1.000	1.260	1.000	1.235	0.9806
Firm	2	B	1.285	1.310	0.9806	1.224	0.8171	1.000	0.8012	0.9283	1.077	1.000	1.056	0.9806
Firm	3	C	0.8909	1.310	0.6799	1.000	0.6934	1.000	0.6799	0.6934	1.000	1.000	0.9806	0.9806
Firm	4	D	0.8719	1.310	0.6654	1.027	0.6607	1.000	0.6479	0.6525	1.040	1.000	1.020	0.9806
Firm	5	E	0.7985	1.310	0.6094	1.000	0.6214	1.000	0.6094	0.5848	1.063	1.000	1.042	0.9806
Year	2		2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Year	3		1.084	1.000	1.084	0.8693	1.247	0.7402	1.685	1.122	0.9666	1.000	0.9666	1.000
Year	4		0.5372	1.125	0.4775	1.320	0.3839	1.351	0.2679	0.3839	1.320	1.000	1.244	0.9428
Mean			1.052	1.310	0.8030	1.047	0.7823	1.000	0.7671	0.7552	1.084	1.000	1.063	0.9806
Minim			0.2652	1.000	0.2357	0.5000	0.2000	0.5000	0.1257	0.2000	0.6000	1.000	0.6000	0.9428
Maxim			2.121	2.000	2.000	2.000	2.000	2.000	2.182	1.600	2.000	1.000	1.886	1.000

Table 4. MEASURES OF TECHNICAL, SCALE and MIX EFFICIENCY (ASSUMING CRS)

Obs	Year	Firm	OTE (D)	OSE (E)	OME (F)	ITE (H)	ISE (I)	IME (J)
1	1	1	0.5000	1.000	1.000	0.5000	1.000	1.000
2	1	2	0.5000	1.000	1.000	0.5000	1.000	1.000
3	1	3	1.000	1.000	1.000	1.000	1.000	1.000
4	1	4	0.8000	1.000	1.000	0.8000	1.000	1.000
5	1	5	0.8333	1.000	1.000	0.8333	1.000	1.000
6	2	1	0.5000	1.000	1.000	0.5000	1.000	1.000
7	2	2	0.5000	1.000	1.000	0.5000	1.000	1.000
8	2	3	1.000	1.000	1.000	1.000	1.000	1.000
9	2	4	0.8000	1.000	1.000	0.8000	1.000	1.000
10	2	5	0.8333	1.000	1.000	0.8333	1.000	1.000
11	3	1	0.5000	1.000	1.000	0.5000	1.000	1.000
12	3	2	1.000	1.000	0.5000	1.000	1.000	1.000
13	3	3	1.000	1.000	1.000	1.000	1.000	1.000
14	3	4	1.000	1.000	0.6667	1.000	1.000	1.000
15	3	5	0.5000	1.000	0.6667	0.5000	1.000	1.000
16	4	1	1.000	1.000	1.000	1.000	1.000	1.000
17	4	2	0.5000	1.000	1.000	0.5000	1.000	1.000
18	4	3	0.3333	1.000	1.000	0.3333	1.000	1.000
19	4	4	0.2500	1.000	1.000	0.2500	1.000	1.000
20	4	5	0.2000	1.000	1.000	0.2000	1.000	1.000
Firm	1	A	0.5946	1.000	1.000	0.5946	1.000	1.000
Firm	2	B	0.5946	1.000	0.8409	0.5946	1.000	1.000
Firm	3	C	0.7598	1.000	1.000	0.7598	1.000	1.000
Firm	4	D	0.6325	1.000	0.9036	0.6325	1.000	1.000
Firm	5	E	0.5133	1.000	0.9036	0.5133	1.000	1.000
Year	1		0.6988	1.000	1.000	0.6988	1.000	1.000
Year	2		0.6988	1.000	1.000	0.6988	1.000	1.000
Year	3		0.7579	1.000	0.7402	0.7579	1.000	1.000
Year	4		0.3839	1.000	1.000	0.3839	1.000	1.000
Mean			0.6139	1.000	0.9276	0.6139	1.000	1.000
Minim			0.2000	1.000	0.5000	0.2000	1.000	1.000
Maxim			1.000	1.000	1.000	1.000	1.000	1.000

Table 5. DECOMPOSITION OF MOORSTEEN-BJUREK TFP INDEX (ASSUMING CRS)

Obs	Year	Firm	TFP Index = (A)	Tech Change x (B)	Eff Change (C)	dOTE (D)	dOSE (E)	dOME (F)	dROSE (G)	dITE (H)	dISE (I)	dIME (J)	dRISE (K)	dRME (L)
						Components of Efficiency Change								
6	2	1	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	2	2	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	2	3	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	2	4	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	2	5	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	3	1	0.8165	1.000	0.8165	1.000	1.000	1.000	0.8165	1.000	1.000	1.000	0.8165	0.8165
12	3	2	2.309	1.000	2.309	2.000	1.000	0.5000	2.309	2.000	1.000	1.000	1.155	1.155
13	3	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	3	4	1.614	1.000	1.614	1.250	1.000	0.6667	1.936	1.250	1.000	1.000	1.291	1.291
15	3	5	0.8485	1.000	0.8485	0.6000	1.000	0.6667	2.121	0.6000	1.000	1.000	1.414	1.414
16	4	1	3.674	1.125	3.266	2.000	1.000	1.000	1.633	2.000	1.000	1.000	1.633	1.633
17	4	2	0.6495	1.125	0.5774	0.5000	1.000	2.000	0.5774	0.5000	1.000	1.000	1.155	1.155
18	4	3	0.3536	1.125	0.3143	0.3333	1.000	1.000	0.9428	0.3333	1.000	1.000	0.9428	0.9428
19	4	4	0.2296	1.125	0.2041	0.2500	1.000	1.500	0.5443	0.2500	1.000	1.000	0.8165	0.8165
20	4	5	0.3286	1.125	0.2921	0.4000	1.000	1.500	0.4869	0.4000	1.000	1.000	0.7303	0.7303
Firm	1	A	1.817	1.310	1.387	1.260	1.000	1.000	1.101	1.260	1.000	1.000	1.101	1.101
Firm	2	B	1.442	1.310	1.101	1.000	1.000	1.000	1.101	1.000	1.000	1.000	1.101	1.101
Firm	3	C	0.8909	1.310	0.6799	0.6934	1.000	1.000	0.9806	0.6934	1.000	1.000	0.9806	0.9806
Firm	4	D	0.9050	1.310	0.6906	0.6786	1.000	1.000	1.018	0.6786	1.000	1.000	1.018	1.018
Firm	5	E	0.8231	1.310	0.6282	0.6214	1.000	1.000	1.011	0.6214	1.000	1.000	1.011	1.011
Year	2		2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Year	3		1.209	1.000	1.209	1.084	1.000	0.7402	1.506	1.084	1.000	1.000	1.115	1.115
Year	4		0.5765	1.125	0.5124	0.5065	1.000	1.351	0.7489	0.5065	1.000	1.000	1.012	1.012
Mean			1.117	1.310	0.8525	0.8190	1.000	1.000	1.041	0.8190	1.000	1.000	1.041	1.041
Minim			0.2296	1.000	0.2041	0.2500	1.000	0.5000	0.4869	0.2500	1.000	1.000	0.7303	0.7303
Maxim			3.674	2.000	3.266	2.000	1.000	2.000	2.309	2.000	1.000	1.000	1.633	1.633

Table 6. DECOMPOSITION OF MALMQUIST TFP INDEX

						Fare et al (1994) Components of Efficiency Change			
Obs	Year	Firm	TFP Index = (A)	Tech Change x (B)	Eff Change (C)	Pure OTE x (D)	Scale Eff (E)	Pure ITE x (H)	Scale Eff (I)
6	2	1	2.000	2.000	1.000	1.000	1.000	1.000	1.000
7	2	2	2.000	2.000	1.000	1.000	1.000	1.000	1.000
8	2	3	2.000	2.000	1.000	1.000	1.000	1.000	1.000
9	2	4	2.000	2.000	1.000	1.000	1.000	1.000	1.000
10	2	5	2.000	2.000	1.000	1.000	1.000	1.000	1.000
11	3	1	1.000	1.000	1.000	0.5000	2.000	1.000	1.000
12	3	2	2.000	1.000	2.000	1.833	1.091	1.600	1.250
13	3	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	3	4	1.250	1.000	1.250	1.083	1.154	1.111	1.125
15	3	5	0.6000	1.000	0.6000	0.5000	1.200	1.000	0.6000
16	4	1	2.121	1.061	2.000	2.000	1.000	1.000	2.000
17	4	2	0.5303	1.061	0.5000	1.000	0.5000	0.5000	1.000
18	4	3	0.3536	1.061	0.3333	1.000	0.3333	0.3333	1.000
19	4	4	0.2652	1.061	0.2500	1.000	0.2500	0.2500	1.000
20	4	5	0.4243	1.061	0.4000	2.000	0.2000	0.2000	2.000
Firm	1	A	1.619	1.285	1.260	1.000	1.260	1.000	1.260
Firm	2	B	1.285	1.285	1.000	1.224	0.8171	0.9283	1.077
Firm	3	C	0.8909	1.285	0.6934	1.000	0.6934	0.6934	1.000
Firm	4	D	0.8719	1.285	0.6786	1.027	0.6607	0.6525	1.040
Firm	5	E	0.7985	1.285	0.6214	1.000	0.6214	0.5848	1.063
Year	2		2.000	2.000	1.000	1.000	1.000	1.000	1.000
Year	3		1.084	1.000	1.084	0.8693	1.247	1.122	0.9666
Year	4		0.5372	1.061	0.5065	1.320	0.3839	0.3839	1.320
Mean			1.052	1.285	0.8190	1.047	0.7823	0.7552	1.084
Minim			0.2652	1.000	0.2500	0.5000	0.2000	0.2000	0.6000
Minim			2.121	2.000	2.000	2.000	2.000	1.600	2.000

REFERENCES

Afriat, S.N. (1972) 'The Theory of International Comparisons of Real Income and Prices', in Daly, D.J. (ed.), *International Comparisons of Prices and Output*, Columbia University Press, New York.

Balk, B.M. (1995) 'Axiomatic Price Index Theory: A Survey', *International Statistical Review*, 63 (1): 69-93.

Balk, B.M. (1998) *Industrial Price, Quantity, and Productivity Indices: The Micro-Economic Theory and an Application*, 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.

Balk, B.M. (2005) 'The Many Decompositions of Productivity Change', Erasmus University, Rotterdam.

Bennet, T.L. (1920) 'The Theory of Measurement of Changes in Cost of Living', *Journal of the Royal Statistical Society*, 83 (3): 455-462.

Bjurek, H. (1996) 'The Malmquist Total Factor Productivity Index', *Scandinavian Journal of Economics*, 98 (2): 303-313.

Byushgens, S.S. (1925) 'Ob odnom klasse giperpoverkhnostey: po povodu 'idealnovo indeksa' Irving Fisher, a pokupatelnoi sili deneg ', *Mathematisheskii Sbornik* 32: 625-631.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982a) 'The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity', *Econometrica*, 50 (6): 1393-1414.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982b) 'Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers', *The Economic Journal*, 92 (365): 73-86.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J. and Battese, G.E. (2005) *An Introduction to Efficiency and Productivity Analysis*, 2nd edn., New York: Springer.

Copeland, M.A. (1937) 'Concepts of National Income', *Studies in Income and Wealth*, National Bureau of Economic Research, New York.

Diewert, W.E. (1976) 'Exact and Superlative Index Numbers', *Journal of Econometrics*, 4 (2): 115-145.

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

Eichhorn, W. (1978) *Functional Equations in Economics*, Reading, Massachusetts: Addison-Wesley.

Fare, R., Grosskopf, S., Lindgren, B. and Roos, P. (1992) 'Productivity Changes in Swedish Pharmacies 1980-1989: A Non-Parametric Malmquist Approach', *Journal of Productivity Analysis*, 3: 85-101.

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

Fare, R., Grosskopf, S., Norris, M. and Zhang, Z. (1994) 'Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries', *American Economic Review*, 84 (1): 66-83.

Fare, R., Grosskopf, S. and Roos, P. (1996) 'On Two Definitions of Productivity', *Economics Letters*, 55: 169-274.

Fare, R., Grosskopf, S. and Roos, P. (1998) 'Malmquist Productivity Indexes: A Survey of Theory and Practice', in Fare, R., Grosskopf, S. and Russell, R.R. (eds.), *Index Numbers: Essays in Honour of Sten Malmquist*, Kluwer Academic Publishers, Boston.

Fare, R. and Primont, D. (1994) 'The Unification of Ronald W. Shephard's Duality Theory', *Journal of Economics*, 60 (2): 199-207.

Farrell, M.J. (1957) 'The Measurement of Productive Efficiency', *Journal of the Royal Statistical Society, Series A (General)*, 120 (3): 253-290.

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

Frisch, R. (1936) 'Annual Survey of General Economic Theory: The Problem of Index Numbers', *Econometrica*, 4 (1): 1-38.

Good, D., Nadiri, M.I. and Sickles, R. (1997) 'Index Number and Factor Demand Approaches to the Estimation of Productivity', in Pesaran, H. and Schmidt, P. (eds.), *Handbook of Applied Econometrics*, Basil Blackwell, Oxford.

Grifell-Tatje, E. and Lovell, C.A.K. (1995) 'A Note on the Malmquist Productivity Index', *Economics Letters*, 47 (2): 169-175.

Grifell-Tatje, E. and Lovell, C.A.K. (1999) 'Profits and Productivity', *Management Science*, 45 (9): 1177-1193.

Hicks, J.R. (1961) 'Measurement of Capital in Relation to the Measurement of Other Economic Aggregates', in Lutz, F.A. and Hague, D.C. (eds.), *The Theory of Capital*, Macmillan, London.

Hulten, C.R. (1973) 'Divisia Index Numbers', *Econometrica*, 41 (6): 1017-1025.

Jorgenson, D.W. and Grilliches, Z. (1967) 'The Explanation of Productivity Change', *The Review of Economic Studies*, 34 (3): 249-283.

Konus, A.A. (1924) 'The Problem of the True Index of the Cost of Living', *Econometrica*, 7 (1 (Jan 1939)): 10-29.

Konus, A.A. and Byushgens, S.S. (1926) 'K probleme pokupatelnoi cili deneg (English translation of Russian title: On the Problem of the Purchasing Power of Money)', *Voprosi Konyunkturi* II (1): 151-171.

Kuosmanen, T. and Sipiläinen, T. (2009) 'Exact decomposition of the Fisher ideal total factor productivity index', *Journal of Productivity Analysis*, forthcoming.

Lovell, C.A.K. (2003) 'The Decomposition of Malmquist Productivity Indexes', *Journal of Productivity Analysis*, 20: 437-458.

Luenberger, D.G. (1992) 'Benefit Functions and Duality', *Journal of Mathematical Economics*, 21: 461-481.

Malmquist, S. (1953) 'Index Numbers and Indifference Surfaces', *Trabajos de Estatistica*, 4: 209-242.

Moorsteen, R.H. (1961) 'On Measuring Productive Potential and Relative Efficiency', *The Quarterly Journal of Economics*, 75 (3): 151-167.

Nadiri, M.I. and Nandi, B. (1999) 'Technical Change, Markup, Divestiture, and Productivity Growth in the U.S. Telecommunications Industry', *Review of Economics and Statistics*, 81 (3): 488-498.

O'Donnell, C.J. (2009) 'Measuring and Decomposing Agricultural Productivity and Profitability Change', *53rd Annual Conference of the Australian Agricultural and Resource Economics Society*, Cairns, Australia.

Pigou, A.C. (1912) *Wealth and Welfare*, London: Macmillan.

Pollack, R.A. (1971) 'The Theory of the Cost of Living Index', *Office of Prices and Living Conditions Research Discussion Papers*, Bureau of Labor Statistics, Washington, D.C.

Ray, S.C. and Mukherjee, K. (1996) 'Decomposition of the Fisher Ideal Index of Productivity: A Non-Parametric Dual Analysis of US Airlines Data', *The Economic Journal*, 106 (439): 1659-1678.

Richter, M.K. (1966) 'Invariance Axioms and Economic Indexes', *Econometrica*, 34 (4): 739-755.

Sahoo, B.K. and Tone, K. (2008) 'Radial and non-radial decompositions of profit change: With an application to Indian banking', *European Journal of Operational Research*, xxx (xxx): xxx-xxx.

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

Shephard, R.W. (1970) *The Theory of Cost and Production Functions*, Princeton: Princeton University Press.

Solow, R.M. (1957) 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, 39 (3): 312-320.

Wald, A. (1939) 'A New Formula for the Index of Cost of Living', *Econometrica*, 7: 319-331.