Decomposing Bjurek Productivity Indexes into Explanatory Factors

June 24, 2014

Erwin Diewert,[1]
Discussion Paper 14-07,
School of Economics,
University of British Columbia,
Vancouver, B.C.,
Canada, V6N 1Z1.
Email: erwin.diewert@ubc.ca

Kevin J. Fox,
School of Economics,
University of New South Wales,
Sydney 2052,
Australia.
Email : K.Fox@unsw.edu.au

Abstract

Caves, Christensen, Diewert introduced Malmquist output, input and productivity indexes into production theory in a systematic way. This paper revisits the debate on how to decompose Bjurek's concept of a Malmquist productivity index into explanatory factors, with a focus on extracting technical progress, technical efficiency change, and returns to scale components. In order to define these components, a reference technology is required. The paper does not make any convexity assumptions on the reference technology but instead follows the example of Tulkens and his coauthors in assuming that the reference technology satisfies free disposability assumptions. The existence and properties of the underlying distance functions of the productivity decomposition are proven under relatively unrestrictive assumptions. The paper provides for the first time a theoretical justification for the geometric average form of the Bjurek productivity index.

Journal of Economic Literature Numbers

C43, D24, E23

Key Words

Productivity indexes, Malmquist Moorsteen Bjurek indexes, technical efficiency, technical progress, returns to scale, Data Envelopment Analysis, Free Disposal Hulls, nonparametric approaches to production theory, distance functions.

[1] The authors gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada and the Australian Research Council (LP0884095). The authors also thank Bert Balk, Rolf Färe, Shawna Grosskopf, Knox Lovell and two referees for helpful comments. Contact author: Kevin Fox, School of Economics & Centre for Applied Economic Research, University of New South Wales, Sydney 2052 Australia, K.Fox@unsw.edu.au

1. Introduction

Malmquist input, output and productivity indexes, initially defined by Caves, Christensen and Diewert (1982) (CCD), have attracted much interest in the literature on productivity analysis. These theoretical indexes use distance functions to represent the technology. CCD proposed a method for estimating a theoretical Malmquist productivity index for a firm whose technology could be represented by a translog technology. The assumptions CCD used were relaxed by Diewert and Fox (2010), who assumed that the firm solved a monopolistic profit maximization problem when the technology exhibited increasing returns to scale. However, Diewert and Fox continued to assume a translog technology whereas in the present paper, we will take a nonparametric approach to the measurement of productivity change.

Since the contribution of CCD, and particularly following Färe, Grosskopf and Lovell (1994) and Bjurek (1996), an alternative form of the Malmquist productivity index has been proposed and there have been attempts to decompose this alternative Malmquist productivity index into technical change, efficiency change and scale change components using the linear programming based Data Envelopment Analysis (DEA) approach for estimating various Malmquist indexes.[2] Determining the appropriate way to do this has led to a significant debate.[3]

This paper revisits this debate of how to decompose the Bjurek form of the Malmquist productivity index into explanatory factors, with a focus on extracting technical progress, technical efficiency change and returns to scale components. In particular, our paper largely follows the approach taken by Balk (2001), except that we do not make use of his cone technology, which is assumed to envelop the actual technology.[4] Furthermore, our regularity conditions on the reference technology are fairly weak and do not make any convexity assumptions. We follow the approach of Tulkens (1993) and his coauthors and replace convexity assumptions by free disposability assumptions.

Our approach is consistent with what O'Donnell (2012; 255) describes as a "bottom-up approach", which starts with basic definitions of the components of productivity and then uses these components to form a multiplicative decomposition of the Bjurek productivity index into the previously defined explanatory factors. However, given that we are not assuming convexity, we start even further back than specifying expressions for the components. Instead we start with proving the existence and the properties of the distance

[2] See Afriat (1972), Hanoch and Rothschild (1972), Charnes, Cooper and Rhodes (1978), Diewert and Parkan (1983), Varian (1984) and Charnes and Cooper (1985) on early applications of DEA.

[3] See Bjurek (1996), Ray and Desli (1997), Balk (2001), Lovell (2003), Nemoto and Goto (2005), O'Donnell (2010) (2012), Briec and Kirstens (2011) and Kerstens and Van de Woestyne (2014). None of these authors work with our weak regularity conditions nor do they derive our decomposition. Grosskopf (2003) provides additional references and a nice historical review of the early stages of the debate.

[4] The enveloping cone technology approach runs into difficulties if the reference technology exhibits global increasing returns.

functions which form these components under our relatively unrestrictive regularity conditions

In deriving our productivity decomposition, the paper makes a number of additional contributions to the existing literature, including resolving some fundamental issues without resorting to specific functional form assumptions. For example, a previously unresolved issue in the literature was that Bjurek (1996; 310) provided no justification for taking the geometric mean of the alternative indexes to form his productivity index, and to our best knowledge neither have subsequent authors.[5] We show that, besides satisfying other properties from index number theory, the geometric Bjurek index is the *only* index that is a homogeneous symmetric average of Laspeyres and Paasche Bjurek productivity indexes that satisfies the desirable time reversal property. The same property is shown for each of the components of the decomposition of the Bjurek index. Hence, for the first time there is a solid theoretical justification for the geometric form of the Bjurek index, as well as its components.

Differentiability is frequently assumed in the literature in order to get a *local* returns to scale term; e.g., differentiability is assumed in the productivity decompositions of Caves, Christensen and Diewert (1982; 1404-1408), Balk (2001), Nemoto and Goto (2005), Diewert and Fox (2010) and O'Donnell (2010) (2012). We define measures of *global* returns to scale term and show that these measures have sensible properties without relying on differentiability assumptions.

In sections 2, 3 and 4 below, we follow Bjurek (1996) and define Malmquist input, output and productivity indexes for a production unit when knowledge of the reference *best practice technology* is available for the unit for the two periods under consideration.[6] Some of the axiomatic properties of these indexes are developed in these sections.

In sections 5, 6 and 7 below, we follow Färe, Grosskopf and Lovell (1994), Balk (2001), Nemoto and Goto (2005) and O'Donnell (2012) for the most part and define measures of technical efficiency, technical change and returns to scale for a production unit using distance functions and assuming knowledge of the best practice technology for the two periods under consideration. Finally, in section 8 we use the measures defined in the previous sections to decompose the productivity of a production unit into the product of three explanatory factors. Section 9 concludes. The Appendix lists our assumptions on the reference technology and develops the properties of the associated distance functions under our relatively unrestrictive assumptions.

2. Malmquist Input Indexes

CCD used the distance function method for representing a technology in order to define families of input, output and productivity indexes. The distance function was introduced into the economics literature in the consumer context by Malmquist (1953) and in the

[5] Balk (2001; 171) justifies the use of the geometric mean simply as a way of avoiding a choice between alternative specifications. See also Nemoto and Goto (2005; 619-620) and O'Donnell (2012; 257-258).
[6] The term "reference technology" originates with Grosskopf (1986).

production context by Shephard (1953). The CCD definitions for Malmquist output and input indexes were generalized by Bjurek (1996) to cover applications of these indexes when estimates of best practice technologies are available. In this section and the following one, we give the basic theoretical definitions for the CCD-Bjurek *input* and *output indexes*. These input and output indexes are then used in order to define a family of (Hicks-Moorsteen) Bjurek *productivity indexes*.[7]

Let S^t be a *reference production possibilities set* for a production unit for periods t = 0,1. This reference technology could be determined via a data envelopment application or could be estimated using econometric techniques. It represents the *best practice* or *efficient technology set* for period t. We assume that S^t is a nonempty closed subset of the nonnegative orthant in Euclidean M+N dimensional space. If (y,x) belongs to S^t, then the nonnegative vector of M outputs $y \equiv [y_1,...,y_M] \geq 0_M$ can be produced using the period t technology by the vector of N nonnegative inputs $x \equiv [x_1,...,x_N] \geq 0_N$.[8]

Using the period t reference technology set S^t and given a nonnegative, nonzero output vector $y > 0_M$ and a strictly positive input vector $x >> 0_N$, the period t *input distance function* D^t for periods t = 0,1 can be defined as follows:

(1) $D^t(y,x) \equiv \max_{\delta>0} \{\delta: (y,x/\delta)\in S^t\}$.

Thus given the nonnegative, nonzero vector of outputs y and the strictly positive vector of inputs x, $D^t(y,x)$ is the maximal amount that the input vector x can be deflated so that the deflated input vector $x/D^t(y,x)$ can produce the vector of outputs y using the period t technology S^t.

Instead of deflating the input vector x so that the resulting deflated vector is just big enough to produce the vector of outputs y, we could think of deflating the output vector so that the resulting deflated output vector is just producible by the input vector x. Thus given $y > 0_M$ and $x >> 0_N$ and the period t reference technology S^t, the period t *output distance function* d^t for periods t = 0,1 can be defined as follows:

(2) $d^t(y,x) \equiv \min_{\delta>0} \{\delta: (y/\delta,x)\in S^t\}$.

[7] CCD (1982; 1402) provided a definition for a Malmquist productivity index but their definition can best be interpreted as an index of technical progress. An alternative productivity index is the Bjurek productivity index and we will use this concept in the present paper. This second concept was defined initially by Moorsteen (1961) (for the case of two outputs and two inputs). Diewert (1992; 240) provided a general definition for this second type of productivity index in terms of distance functions but assumed technical efficiency (as did Moorsteen). Bjurek (1996) generalized these definitions to cover the case of technical inefficiency. Diewert (1992; 240) attributed this second productivity index concept to Hicks (1961; 22) (for his apparent description of the concept in words) and Moorsteen (1961) (for the case of two outputs and two inputs), and hence termed it the “Hicks-Moorsteen approach to productivity indexes”. The subsequent description by Hicks (1981; 253-265) of his “opportunity cost” methodology reveals that he did have a distance function interpretation for his output and input quantity indexes but his description was not very clear. Färe, Grosskopf and Roos (1996) provided conditions for the equivalence of the two types of Malmquist productivity indexes.

[8] Notation: $y \geq 0_M$ means each component of the vector y is nonnegative, $y >> 0_M$ means that each component is strictly positive, and $y > 0_M$ means $y \geq 0_M$ but $y \neq 0_M$.

It is not immediately clear that the maximum in (1) or the minimum in (2) will exist. In fact, in order to obtain the existence of the functions D^t and d^t defined by (1) and (2), some restrictions on the production possibilities sets S^t are required (in addition to the assumption that S^t is a closed, nonempty subset of the nonnegative orthant). Below, we postulate a simple set of restrictions on the S^t, properties P1–P7, which will guarantee the existence of these input and output distance functions.

Given a reference output vector $y > 0_M$ and two strictly positive input vectors $x^0 >> 0_N$ and $x^1 >> 0_N$, the input distance function $D^t(y,x)$ that corresponds to the period t reference technology S^t can be used to define the following family of *Malmquist input indexes*,[9] $Q(x^0,x^1,y,t)$:

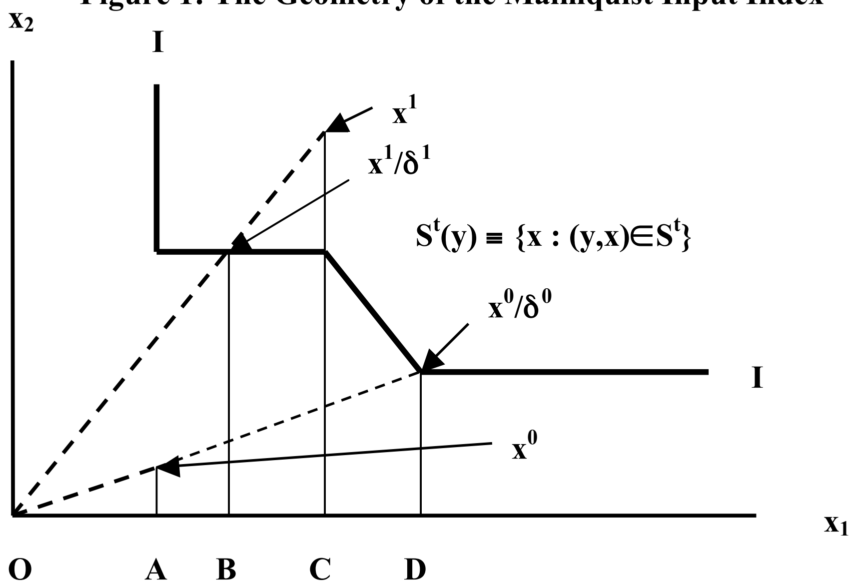
(5) $Q(x^0,x^1,y,S^t) \equiv D^t(y,x^1)/D^t(y,x^0)$.

A value of the index greater than one implies that the input vector x^1 is larger than the input vector x^0, using y as a reference output and the period t best practice technology, S^t, as the reference technology. In the following sections, x^0 will be interpreted as the input vector that corresponds to a production unit that operates in period 0 and x^1 will be interpreted as the input vector that corresponds to a production unit that operates in period 1. If N = 1, so that there is only one input, then $Q(x_1^0,x_1^1,y,S^t)$ equals x_1^1/x_1^0.[10] The geometry of the Malmquist input index for two inputs is illustrated in Figure 1.

[9] The use of input distance functions to define input indexes can be traced back to Moorsteen (1961; 462). Fisher and Shell (1972; 51), Diewert (1980; 462) and Caves, Christensen and Diewert (1982; 1396) all used variants of this concept in the context of production theory. The general definition of the input index given by (5) is due to Bjurek (1996; 307).

[10] Let N = 1 and let $y > 0_M$, $x_1^0 > 0$ and $x_1^1 > 0$. Let S^t satisfy the regularity conditions P1-P4 to be introduced below. Then it can be verified that $(x: (y,x) \in S^t)$ is the set $\{x_1 : x_1 \geq g(y) > 0\}$ where g(y) is the minimum amount of input required to produce the vector of outputs y using the technology set S^t. Thus $D^t(y,x_1^0) = \max_\delta \{\delta : (y,x_1^0/\delta) \in S^t\} = \max_\delta \{\delta : x_1^0/\delta \geq g(y)\} = \delta^0$ where $\delta^0 = x_1^0/g(y) > 0$. Similarly $D^t(y,x_1^1) = x_1^1/g(y) > 0$. Thus $Q(x_1^0,x_1^1,y,S^t) \equiv D^t(y,x_1^1)/D^t(y,x_1^0) = x_1^1/x_1^0$.

Figure 1: The Geometry of the Malmquist Input Index



Given a reference technology set S^t and a reference output vector $y > 0_M$, the set of inputs x that can produce the vector of outputs y is $S^t(y) \equiv \{x : (y,x) \in S^t\}$. In Figure 1, this set of feasible inputs lies on and above the kinked boundary line I-I. Note that the period 1 input vector $x^1 \equiv [x_1{}^1,x_2{}^1]$ lies in the interior of $S^t(y)$ while the period 0 input vector $x^0 \equiv [x_1{}^0,x_2{}^0]$ is exterior to $S^t(y)$. Define $\delta^0 \equiv D^t(y,x^0)$ so that x^0/δ^0 is on the boundary line I-I. It can be seen that δ^0 is less than one and δ^0 equals OA/OD, the distance OA divided by the distance OD. Define $\delta^1 \equiv D^t(y,x^1)$ so that x^1/δ^1 is on the boundary line I-I. It can be seen that δ^1 is greater than one and, δ^1 equals OC/OB. Thus the input index $Q(x^0,x^1)$ is equal to [OC/OB]/[OA/OD] = [OC/OB][OD/OA] where the distance ratios OC/OB and OD/OA are both greater than one in this case. It can be seen that if both input vectors x^0 and x^1 are on the frontier of the input production possibilities set $S^t(y)$ (i.e., they are both on the boundary line I-I), then $Q(x^0,x^1)$ equals one and the input vectors are regarded as having equivalent size.[11] If x^0 is below the boundary line I-I and x^1 is on the boundary line or above it, then $Q(x^0,x^1)$ is greater than one and x^0 is regarded as being a smaller amount of aggregate input than the amount represented by x^1. This is the idea behind the Malmquist (1953) index, which was originally developed in the consumer context.[12]

[11] Note that the Malmquist input quantity index has no separate role for *mix effects*; i.e., there is no separate role for the possibility that x^1 may not be proportional to x^0. The role of mix effects will be revisited when considering our returns to scale measure in section 7.

[12] Note that if x is on the boundary line I-I, then (y,x) is on the boundary of S^t. However, this does not necessarily mean that (y,x) is *efficient* for the period t technology. Tulkens (1993; 192) recognized this problem with radial efficiency measures in the context of free disposal type reference technologies and he commented on the problem as follows: “As no *radial* measure can circumvent this [difficulty], and in view of the easy interpretability of radial measures, it seems preferable to use them in numerical applications, with slacks reported separately.” For a comprehensive discussions on the difficulties associated with measuring technical efficiency, see Russell and Schworm (2009) (2011).

Now suppose that the strictly positive vector x^0 in Figure 1 is shifted down to the point A on the x_1 axis. It can be seen that in this case, $D^t(y,x^0)$ is not well defined; i.e., the x_1 axis never touches the input production possibilities set $S^t(y)$. Thus the restriction that the vectors x^0 and x^1 be strictly positive ensures that the Malmquist input index is well defined. The example in Figure 1 is also consistent with the *free disposability of inputs* and this is a restriction on the technology that we will impose that will ensure that the input distance functions are well defined for all positive input vectors.

We will now list our regularity conditions on the reference technology set S^t that will ensure that the input distance function $D^t(y,x)$ is well defined. Suppose that the reference technology set S^t satisfies the following regularity conditions:[13]

P1. S is a nonempty closed subset of the nonnegative orthant in Euclidean M+N dimensional space.
P2. For every $y \geq 0_M$, there exists an $x \geq 0_N$ such that $(y,x)\in S$.
P3. $(y,x^1)\in S$, $x^2 \geq x^1$ implies $(y,x^2)\in S$.
P4. $y > 0_M$ implies that $(y,0_N)\notin S$.

The interpretation of P2 is that every finite output vector y is producible by a finite input vector x, if S satisfies P3 then there is free disposability of inputs, and P4 says that zero amounts of all inputs cannot produce a positive output; i.e., there is no free lunch in production.

In the Appendix we show that $D^t(y,x)$ then satisfies the following regularity conditions with respect to x over the positive orthant, $\Omega_N \equiv \{x: x >> 0_N\}$ in N dimensional space: for $y > 0_M$, $D^t(y,x)$ is *positive*, (positively) *linearly homogeneous*, *nondecreasing* (*increasing if all inputs increase*) and *continuous function* of x over Ω_N.

Let S^t satisfy properties P1–P4 and let $y > 0_M$. We now look at the axiomatic properties of $Q(x^0,x^1,y,S^t)$ defined by (5) above with respect to the two input vectors, x^0 and x^1. For brevity, we denote $Q(x^0,x^1,y,S^t)$ by $Q(x^0,x^1)$. Using the properties of the input distance function $D^t(y,x)$ listed in the paragraph above, it is reasonably straightforward to show that $Q(x^0,x^1)$ satisfies the following 12 properties for $x^0 = [x_1^0,...,x_N^0] >> 0_N$ and $x^1 = [x_1^1,...,x_N^1] >> 0_N$:

A1. *Identity*: $Q(x,x) = 1$; i.e., if the period 0 and 1 quantity vectors are equal to $x >> 0_N$, then the index is equal to unity.

A2: *Weak Monotonicity in Current Period Quantities*: $Q(x^0,x^1) \leq Q(x^0,x)$ if $x^1 < x$; i.e., if any period 1 quantity increases, then the quantity index increases or remains constant.

[13] For discussions on regularity conditions, see Färe and Lovell (1978), Deprins, Simar and Tulkens (1984), Färe, Grosskopf and Lovell (1985), Grosskopf (1986), Färe (1988), Tulkins (1993), Färe and Primont (1995), Coelli, Rao and Battese (1997), Balk (2001) (2003), O'Donnell (2010) (2012), Briec and Kerstens (2011), Zelenyuk (2013) and Kerstens and Van de Woestyne (2014). As mentioned earlier, none of these authors used our regularity conditions.

A3: *Strong Monotonicity in Current Period Quantities*: $Q(x^0,x^1) < Q(x^0,x)$ if $x^1 \ll x$; i.e., if all period 1 input quantities increase, then the quantity index increases.

A4: *Weak Monotonicity in Base Period Quantities*: $Q(x^0,x^1) \geq Q(x,x^1)$ if $x^0 < x$; i.e., if any period 0 quantity increases, then the quantity index decreases or remains constant.

A5: *Strong Monotonicity in Base Period Quantities*: $Q(x^0,x^1) > Q(x,x^1)$ if $x^0 \ll x$; i.e., if all period 0 input quantities increase, then the quantity index decreases.

A6: *Proportionality in Current Period Quantities*: $Q(x^0,\lambda x^1) = \lambda Q(x^0,x^1)$ if $\lambda > 0$; i.e., if all period 1 quantities are multiplied by the positive number λ, then the resulting quantity index is equal to the initial quantity index multiplied by λ.

A7: *Inverse Proportionality in Base Period Quantities*: $Q(\lambda x^0,x^1) = \lambda^{-1}Q(x^0,x^1)$ if $\lambda > 0$; i.e., if all period 0 quantities are multiplied by the positive number λ, then the resulting quantity index is equal to the initial quantity index multiplied by $1/\lambda$.

A8: *Mean Value*: $\min_n\{x_n^1/x_n^0: n = 1,...,N\} \leq Q(x^0,x^1) \leq \max_n\{x_n^1/x_n^0: n = 1,...,N\}$; i.e., the input quantity index lies between the smallest and largest quantity relatives.[14]

A9: *Time Reversal*: $Q(x^1,x^0) = 1/Q(x^0,x^1)$; i.e., if the data for periods 0 and 1 are interchanged, then the resulting quantity index should equal the reciprocal of the original quantity index.

A10: *Circularity*: $Q(x^0,x^1)Q(x^1,x^2) = Q(x^0,x^2)$; i.e., the quantity index going from period 0 to 1 times the quantity index going from period 1 to 2 equals the quantity index going from period 0 to 2 directly.

The circularity and identity axioms imply time reversal; just set $x^2 = x^0$. Thus circularity is essentially a strengthening of the time reversal property.

A11: *Commensurability*: $Q(\lambda_1 x_1^0,...,\lambda_N x_N^0;\lambda_1 x_1^1,...,\lambda_N x_N^1) = Q(x_1^0,...,x_N^0;x_1^1,...,x_N^1) = Q(x^0,x^1)$ for all $\lambda_1 > 0, ... , \lambda_N > 0$; i.e., if we change the units of measurement for each input, then the input quantity index remains unchanged.

A12: *Continuity*: $Q(x^0,x^1)$ is a jointly continuous function of x^0 and x^1 for $x^0 \gg 0_N$ and $x^1 \gg 0_N$.[15]

[14] Let $\beta \equiv \max_n \{x_n^1/x_n^0: n = 1,...,N\}$. Then $x^1 \leq \beta x^0$ using the positivity of x^0. Thus $Q(x^0,x^1) \leq Q(x^0,\beta x^0)$ (using A2) $= \beta Q(x^0,x^0)$ (using A6) $= \beta$ (using A1). Similarly, let $\alpha \equiv \min_n \{x_n^1/x_n^0: n = 1,...,N\}$. Then $x^1 \geq \alpha x^0$ using the positivity of x^0. Thus $Q(x^0,x^1) \geq Q(x^0,\alpha x^0)$ (using A2) $= \alpha Q(x^0,x^0)$ (using A6) $= \alpha$ (using A1). This proof follows that of Eichhorn (1978; 155) in the price index context.

[15] O'Donnell (2010) (2012) considered many of these axioms and some additional axioms for input quantity indexes. The above axioms are essentially the modification of the axioms used by Diewert (1992) for bilateral price indexes of the form $P(p^0,p^1,q^0,q^1)$ except that Q replaces P, x^0 and x^1 replace p^0 and p^1 and tests involving changes in q^0,q^1 are deleted.

Recall that (5) defines an entire family of Malmquist input quantity indexes, $Q(x^0,x^1,y,S^t)$; i.e., for each reference output vector $y > 0_M$ and for each reference technology set S^t , there is a possibly *different* input quantity index $Q(x^0,x^1,y,S^t)$. The question we now have to address is: if we are comparing the inputs of two different production units who have the observed output and input vectors (y^0,x^0) and (y^1,x^1), what is an appropriate choice of the reference output vector y and the reference technology set S^t to insert into the definition of the Malmquist input quantity index $Q(x^0,x^1,y,S^t)$?

From the viewpoint of the period 0 production unit, the most appropriate choice of a reference output vector y would seem to be the actual output vector produced by the unit in period 0, which is y^0. Similarly, the most appropriate reference technology for the period 0 production unit would appear to be the period 0 best practice technology, S^0. Thus from the viewpoint of the period 0 production unit, the most appropriate input quantity index would appear to be the *Laspeyres type Malmquist input index*, $Q_L(x^0,x^1)$, defined as follows:

(6) $Q_L(x^0,x^1) \equiv Q(x^0,x^1,y^0,S^0) = D^0(y^0,x^1)/D^0(y^0,x^0)$.

Similarly, from the viewpoint of the period 1 production unit, the most appropriate input quantity index is the *Paasche type Malmquist input index*, $Q_P(x^0,x^1)$, which uses the reference output vector y^1 and best practice technology S^1:

(7) $Q_P(x^0,x^1) \equiv Q(x^0,x^1,y^1,S^1) = D^1(y^1,x^1)/D^1(y^1,x^0)$.

Since we have two separate relevant input quantity indexes[16] when comparing the relative size of the input vectors of two production units, it is natural to take a symmetric average of the two indexes defined by (6) and (7) in order to obtain a "final" measure of the relative magnitude of the input vector x^1 relative to x^0. But what form of average should we take? CCD (1982; 1397) found it convenient to take the geometric average of the above two indexes; i.e., define

(8) $Q_{CCD}(x^0,x^1) = [Q_L(x^0,x^1)Q_P(x^0,x^1)]^{1/2}$.

However, CCD chose the geometric average of the Laspeyres and Paasche type Malmquist input indexes because it led to an exact bilateral index number formula when they made various translog assumptions on the underlying technology. In our present context, we want to avoid the use of price information so we need another justification for taking the geometric mean of Q_L and Q_P as opposed to taking some other form of average. The literature following CCD adopted their choice of geometric mean even when not assuming translog technology. Here we examine this choice.

[16] The two input indexes defined by (6) and (7) were the ones that were introduced by Caves, Christensen and Diewert (1982; 1396). Diewert (1992; 235) also endorsed these two input indexes as being "natural" input indexes.

We should choose the form of average strategically so that the resulting index satisfies an important property, which we take to be *Time Reversal*; see A9 above.

At this point, we need a bit of background information on the properties of averages or *means*. Let a and b be two positive numbers. Diewert (1993; 361) defined a *symmetric mean* of a and b as a function m(a,b) that has the following properties:

(9) $m(a,a) = a$ for all $a > 0$ (mean property);
(10) $m(a,b) = m(b,a)$ for all $a > 0$, $b > 0$ (symmetry property);
(11) m(a,b) is a continuous function for $a > 0$, $b > 0$ (continuity property);
(12) m(a,b) is a strictly increasing function in each of its variables (increasingness property).

It can be shown that if m(a,b) satisfies the above properties, then it also satisfies the following property:

(13) $\min\{a,b\} \leq m(a,b) \leq \max\{a,b\}$ (min-max property);

i.e., the mean of a and b, m(a,b), lies between the maximum and minimum of the numbers a and b.[17] Since we have restricted the domain of definition of a and b to be positive numbers, it can be seen that an implication of (13) is that m also satisfies the following property:

(14) $m(a,b) > 0$ for all $a > 0$, $b > 0$ (positivity property).

If in addition, m satisfies the following property, then Diewert (1993) defined m to be a *homogeneous symmetric mean*:

(15) $m(\lambda a,\lambda b) = \lambda m(a,b)$ for all $\lambda > 0$, $a > 0$, $b > 0$.

With the above material on homogeneous, symmetric means in hand, we can prove the following proposition:

Proposition 1: The CCD input quantity index $Q_{CCD}(x^0,x^1)$ defined by (8) above is the *only* index satisfying the time reversal property A9 that is a homogeneous symmetric average of the Laspeyres and Paasche Malmquist input quantity indexes, Q_L and Q_P defined by (6) and (7).

Proof: Assume that the homogeneous mean function m satisfies the positivity and homogeneity properties, (14) and (15) above.

Let $x^0 >> 0_N$ and $x^1 >> 0_N$. Define $a \equiv Q_L(x^0,x^1) > 0$ and $b \equiv Q_P(x^0,x^1) > 0$. Looking at definitions (6) and (7), it can be seen that if we reverse the order of time:

[17] Proof: Let $a \leq b$. Then $a = m(a,a) \leq m(a,b) \leq m(b,b) = b$ using (9), (11) and (12).

(16) $Q_L(x^1,x^0) = 1/a = 1/Q_L(x^0,x^1)$; $Q_P(x^1,x^0) = 1/b = 1/Q_P(x^0,x^1)$.

Define the mean input quantity index Q using the function m as follows:

(17) $Q(x^0,x^1) \equiv m(Q_L(x^0,x^1),Q_P(x^0,x^1)) = m(a,b)$.

where we have used the definitions of the numbers a and b above. For Q to satisfy the time reversal test, the following equation must be satisfied:

$$
\begin{aligned}
(18)\ Q(x^1,x^0) &= m(Q_L(x^1,x^0),Q_P(x^1,x^0)) \\
&= m(a^{-1},b^{-1}) && \text{using (16)} \\
&= 1/Q(x^0,x^1) \\
&= 1/m(a,b) && \text{using (17).}
\end{aligned}
$$

Using the positivity of a and b and property (14) for m, (18) can be rewritten as follows:

$$
\begin{aligned}
(19)\ 1 &= m(a,b)m(b^{-1},a^{-1}) \\
&= am(1,b/a)a^{-1}m(a/b,1) && \text{using property (15) for m} \\
&= m(1,x)m(x^{-1},1) && \text{letting } x \equiv b/a \\
&= m(1,x)x^{-1}m(1,x) && \text{using property (15) for m.}
\end{aligned}
$$

Equation (19) can be rewritten as:

(20) $x = [m(1,x)]^2$.

Take the positive square root of both sides of (20) and obtain

(21) $m(1,x) = x^{1/2}$.

Using property (15) for m again, we have

$$
\begin{aligned}
m(a,b) &= am(1,b/a) \\
&= a[b/a]^{1/2} && \text{using (21)} \\
&= a^{1/2}b^{1/2}.
\end{aligned}
$$

Now substitute for m(a,b) in (17) and we find that $Q(x^0,x^1) = Q_{CCD}(x^0,x^1)$. Q.E.D.[18]

Using the mathematical properties of the input distance functions $D^0(y^0,x)$ and $D^1(y^1,x)$ with respect to the strictly positive input vector x, it is straightforward to establish the following Proposition:

Proposition 2: Let the technology sets S^0 and S^1 satisfy properties P1–P4 and let $y^0 > 0_M$ and $y^1 > 0_M$. Then the CCD Malmquist input quantity index $Q_{CCD}(x^0,x^1)$ defined by (8) above satisfies the axioms A1-A12 listed above for all $x^0 \gg 0_N$ and $x^1 \gg 0_N$.

[18] This proof is a modification of a proof due to Diewert (1997; 138) in the price index context.

The above Proposition implies that $Q_{CCD}(x^0,x^1)$ satisfies the circularity test A10 and this is true but note that this circularity test is conditional on only the two production possibility sets S^0 and S^1 and the two reference output vectors y^0 and y^1. Thus the index $Q_{CCD}(x^0,x^1)$ should be more properly denoted by $Q_{CCD}(x^0,x^1;y^0,y^1;S^0,S^1)$ and the circularity test that Q_{CCD} satisfies is the following one: for all $x^0 >> 0_N$, $x^1 >> 0_N$ and $x^2 >> 0_N$, we have:

(22) $Q_{CCD}(x^0,x^1;y^0,y^1;S^0,S^1)Q_{CCD}(x^1,x^2;y^0,y^1;S^0,S^1) = Q_{CCD}(x^0,x^2;y^0,y^1;S^0,S^1)$.

Thus there is a certain lack of symmetry in the index when input comparisons are made between three or more production units. Hence the CCD Malmquist input index is best suited for bilateral comparisons between a pair of production units (or the same production unit over two time periods) rather than multilateral comparisons between many production units.

We now turn our attention to Malmquist output indexes.

3. Malmquist Output Indexes

Given a strictly positive reference input vector $x >> 0_N$ and two nonnegative, nonzero output vectors $y^0 > 0_M$ and $y^1 > 0_M$, the output distance function $d^t(y,x)$ defined by (2) that corresponds to the period t reference technology S^t can be used to define the following family of *Malmquist output indexes*,[19] $q(y^0,y^1,x,S^t)$:

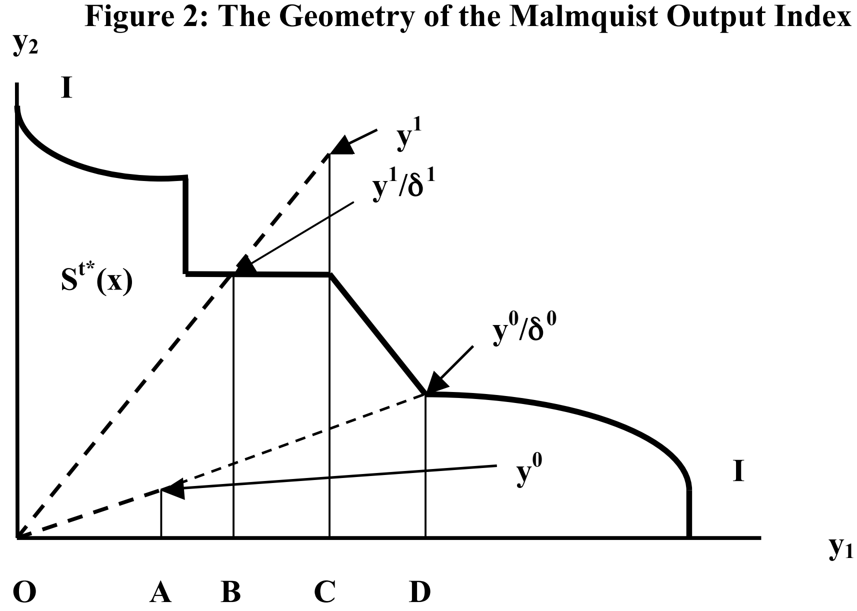
(23) $q(y^0,y^1,x,S^t) \equiv d^t(y^1,x)/d^t(y^0,x)$.

A value of the index greater than one implies that the output vector y^1 is larger than the output vector y^0, using x as a reference output and the period t best practice technology, S^t, as the reference technology. In the following sections, y^0 will be the output vector that corresponds to a production unit that operates in period 0 and y^1 will be the output vector that corresponds to a production unit that operates in period 1. If M = 1, so that there is only one output, then $q(y_1^0, y_1^1,x,S^t)$ equals y_1^1/y_1^0.[20] The geometry of the Malmquist output index for two inputs is illustrated in Figure 2.

[19] The general definition of the output index given by (23) is due to Bjurek (1996; 307). Again Bjurek's definition was a generalization of the CCD definition to allow for technical inefficiency.

[20] Let M = 1 and let $x >> 0_N$, $y^0 > 0$ and $y^1 > 0$. Let S^t satisfy the regularity conditions P1 and P5-P7, to be introduced below. Then it can be verified that $(y : (y,x)\in S^t)$ is the set $\{y : 0 \leq y \leq f(x)\}$ where $f(x) > 0$ is the maximum amount of the single output that can be produced by the strictly positive input vector x using the technology set S^t. Thus $d^t(y^0,x) = \min_\delta \{\delta: (y^0/\delta,x)\in S^t\} = \min_\delta \{\delta: y^0/\delta \leq f(x)\} = \delta^0$ where $\delta^0 = y^0/f(x) > 0$. Similarly $d^t(y^1,x) = y^1/f(x) > 0$. Thus $q(y^0,y^1,x,S^t) \equiv d^t(y^1,x)/d^t(y^0,x) = y^1/y^0$.

Figure 2: The Geometry of the Malmquist Output Index



Given a reference technology set S^t and a reference input vector $x >> 0_N$, the set of outputs y that can be produced by the vector of inputs x is $S^{t*}(x) \equiv \{y : (y,x)\in S^t\}$. In Figure 2, this set of feasible inputs is a subset of the nonnegative orthant and lies on and below the kinked boundary line I-I. Note that the period 1 output vector $y^1 \equiv [y_1^{\,1},y_2^{\,1}]$ lies outside of $S^{t*}(x)$ while the period 0 output vector $y^0 \equiv [y_1^{\,0},y_2^{\,0}]$ is in the interior of $S^{t*}(x)$. Define $\delta^0 \equiv d^t(y^0,x)$ so that y^0/δ^0 is on the boundary line I-I. It can be seen that δ^0 is less than one and δ^0 equals OA/OD, the distance OA divided by the distance OD. Define $\delta^1 \equiv d^t(y^1,x)$ so that y^1/δ^1 is on the boundary line I-I. It can be seen that δ^1 is greater than one and, δ^1 equals OC/OB. Thus the output index $q(x^0,x^1)$ is equal to [OC/OB]/[OA/OD] = [OC/OB][OD/OA] where the distance ratios OC/OB and OD/OA are both greater than one in this case. It can be seen that if both output vectors y^0 and y^1 are on the frontier of the input production possibilities set $S^{t*}(x)$ (i.e., they are both on the boundary line I-I), then $q(y^0,y^1)$ equals one and the output vectors are regarded as having equivalent size. If y^0 is below the boundary line I-I and y^1 is on the boundary line or above it, then $q(y^0,y^1)$ is greater than one and y^0 is regarded as being a smaller amount of aggregate output than the amount represented by y^1. Note that we do not require y^0 and y^1 to be strictly positive vectors in order for the output index to be well defined; we need only $y^0 > 0_M$ and $y^1 > 0_M$.

We will now list our regularity conditions on the reference technology set S^t that will ensure that the output distance function $d^t(y,x)$ is well defined. Suppose that the reference technology set S^t satisfies condition P1 listed in the previous section and the following three additional regularity conditions :

P5. $x \geq 0_N$, $(y,x)\in S$ implies $0_M \leq y \leq b(x)1_M$ where 1_M is a vector of ones of dimension M and $b(x) \geq 0$ is a finite nonnegative bound.
P6. $x >> 0_N$ implies that there exists $y >> 0_M$ such that $(y,x)\in S$.

P7. $(y^1,x) \in S$, $0_M \le y^0 \le y^1$ implies $(y^0,x) \in S$.

The interpretation of P5 is that bounded inputs imply bounded outputs, P6 says that technology is such that every strictly positive input vector can produce a strictly positive vector of outputs, and P7 means that if the input vector x can produce the output vector y^1 and y^0 is equal to or less than y^1, then x can also produce the smaller vector of outputs, y^0 (i.e. free disposability of outputs).

In the Appendix we show that $d^t(y,x)$ satisfies the following regularity conditions with respect to y over the nonnegative orthant excluding the origin, $\Omega_M{}^* \equiv \{y: y > 0_N\}$: for $x >> 0_N$, $d^t(y,x)$ is *positive*, (positively) *linearly homogeneous*, *nondecreasing* (*increasing if all outputs increase*) and *continuous function* of y over $\Omega_M{}^*$.

Let S^t satisfy properties P1 and P5-P7 and let $x >> 0_N$. We now look at the axiomatic properties of $q(y^0,y^1,x,S^t)$ defined by (23) above with respect to the two output vectors, y^0 and y^1. For brevity, we denote $q(y^0,y^1,x,S^t)$ by $q(y^0,y^1)$. Using the properties of the output distance function $d^t(y,x)$ listed in the paragraph above, it is straightforward to show that $q(y^0,y^1)$ satisfies modified versions of properties A1-A12. The modified A1-A11 simply replace $Q(x^0,x^1)$ by $q(y^0,y^1)$ and the strictly positive input quantity vectors x^0 and x^1 by the nonnegative, nonzero output vectors $y^0 > 0_M$ and $y^1 > 0_M$. For example, the first two modified tests are the following ones:

B1. *Identity*: $q(y,y) = 1$; i.e., if the period 0 and 1 quantity vectors are equal to $y > 0_M$, then the output quantity index is equal to unity.

B2: *Weak Monotonicity in Current Period Quantities*: $q(y^0,y^1) \le q(y^0,y)$ if $y^0 > 0_M$ and $0_M < y^1 < y$; i.e., if any period 1 quantity increases, then the quantity index increases or remains constant.

However, the modified test A12 requires that the two output vectors y^0 and y^1 be strictly positive so that the test B12 is the following one:

B12: *Continuity*: $q(y^0,y^1)$ is a jointly continuous function of y^0 and y^1 for $y^0 >> 0_M$ and $y^1 >> 0_M$.

Recall that (23) defines an entire family of Malmquist output quantity indexes, $q(y^0,y^1,x,S^t)$; i.e., for each reference input vector $x >> 0_N$ and for each reference technology set S^t, there is a possibly *different* output quantity index $q(y^0,y^1,x,S^t)$. Thus we now have to address the same type of question that we addressed in the previous section: if we are comparing the outputs of two different production units who have the observed output and input vectors (y^0,x^0) and (y^1,x^1), what is an appropriate choice of the reference input vector x and the reference technology set t to insert into the definition of the Malmquist output quantity index $q(y^0,y^1,x,S^t)$?

From the viewpoint of the period 0 production unit, the most appropriate choice of a reference input vector x would seem to be the actual input vector used by the unit in period 0, which is x^0 (which we assume is a strictly positive vector). Similarly, the most appropriate reference technology for the period 0 production unit would appear to be the period 0 best practice technology, S^0. Thus from the viewpoint of the period 0 production unit, the most appropriate output quantity index would appear to be the *Laspeyres type Malmquist output index*, $q_L(y^0,y^1)$, defined as follows:

$$(24)\ q_L(y^0,y^1) \equiv q(y^0,y^1,x^0,S^0) = d^0(y^1,x^0)/d^0(y^0,x^0).$$

Similarly, from the viewpoint of the period 1 production unit, the most appropriate output quantity index is the *Paasche type Malmquist output index*, $q_P(y^0,y^1)$, which uses the strictly positive reference input vector x^1 and the best practice technology S^1:

$$(25)\ q_P(y^0,y^1) \equiv q(y^0,y^1,x^1,S^1) = d^1(y^1,x^1)/d^1(y^0,x^1).$$

Since we have two separate relevant output quantity indexes[21] when comparing the relative size of the output vectors of two production units, it is natural to take a symmetric average of the two indexes defined by (24) and (25) in order to obtain a "final" measure of the relative magnitude of the output vector y^1 relative to y^0. But what form of average should we take? CCD (1982; 1401) found it convenient to take the geometric average of the above two indexes; i.e., define

$$(26)\ q_{CCD}(y^0,y^1) = [q_L(y^0,y^1)q_P(y^0,y^1)]^{1/2}.$$

The use of the geometric average of q_L and q_P instead of some other form of average can be justified if we want the average Malmquist output index to satisfy the time reversal property B9; i.e., we can establish the following proposition, using the same method of proof as was used in the proof of Proposition 1 in the previous section:

Proposition 3: The CCD output quantity index $q_{CCD}(y^0,y^1)$ defined by (26) above is the *only* index satisfying the time reversal property B9 that is a homogeneous symmetric average of the Laspeyres and Paasche Malmquist output quantity indexes, q_L and q_P defined by (24) and (25).

Assuming that $x >> 0_N$, using the mathematical properties of the output distance functions $d^0(y,x)$ and $d^1(y,x)$ with respect to the nonnegative, nonzero output vector y that are established in the Appendix, it is straightforward to prove the following proposition:

Proposition 4: Let the technology sets S^0 and S^1 satisfy properties P1 and P5–P7 and let $x^0 >> 0_N$ and $x^1 >> 0_N$. Then the CCD Malmquist output quantity index $q_{CCD}(y^0,y^1)$ defined by (26) above satisfies the axioms B1-B11 for all $y^0 > 0_M$ and $y^1 > 0_M$ and B12 for all $y^0 >> 0_M$ and $y^1 >> 0_M$.

The above proposition implies that $q_{CCD}(y^0,y^1)$ satisfies the circularity test B10 and this is true but note that this circularity test is conditional on only the two production possibility

[21] The two output indexes defined by (24) and (25) were the ones that were introduced by Caves, Christensen and Diewert (1982; 1400).

sets S^0 and S^1 and the two reference input vectors x^0 and x^1. Thus the index $q_{CCD}(y^0,y^1)$ is more properly denoted by $q_{CCD}(y^0,y^1,x^0,x^1;S^0,S^1)$ and the circularity test that q_{CCD} satisfies is the following one: for all $y^0 > 0_M$, $y^1 > 0_M$ and $y^2 > 0_M$, we have:

$$(27)\ q_{CCD}(y^0,y^1,x^0,x^1;S^0,S^1)q_{CCD}(y^1,y^2,x^0,x^1;S^0,S^1) = q_{CCD}(y^0,y^2,x^0,x^1;S^0,S^1).$$

Thus as was the case for the CCD Malmquist input index, the CCD Malmquist output index is best suited for bilateral comparisons between a pair of production units (or the same production unit over two time periods) rather than multilateral comparisons between many production units.

We now turn our attention to productivity indexes.

4. Bjurek Productivity Indexes

Having defined families of input and output indexes using distance functions in the previous two sections, it is natural to define a *family of productivity indexes* as the ratio of a family of output indexes to a family of input indexes. Our goal is to compare the productivity of two production units, 0 and 1, in the same industry which have the observed output and input vectors (y^0,x^0) and (y^1,x^1) respectively. We assume that the input vectors are strictly positive, so that $x^0 >> 0_N$ and $x^1 >> 0_N$, and we assume that the output vectors are nonnegative but nonzero so that $y^0 > 0_M$ and $y^1 > 0_M$. Our definition of the productivity index will also utilize a nonnegative, nonzero *reference output vector* $y > 0_M$ and a strictly positive *reference input vector* $x >> 0_N$. Finally, the definition utilizes a *reference technology set* S^t which satisfies the regularity conditions P1–P7. Recall the family of Malmquist input indexes, $Q(x^0,x^1,y,t)$ defined by (5) above, and the family of Malmquist output indexes, $q(y^0,y^1,x,t)$ defined by (23) above. These two families of indexes can be used in order to define the following family of *Bjurek productivity indexes*:

$$(28)\ \Pi(x^0,x^1,y^0,y^1,x,y,S^t) \equiv q(y^0,y^1,x,S^t)/Q(x^0,x^1,y,S^t)$$
$$= [d^t(y^1,x)/d^t(y^0,x)/[D^t(y,x^1)/D^t(y,x^0)].$$

If $\Pi(x^0,x^1,y^0,y^1,x,y,S^t)$ is greater (less) than one, we say that production unit 1 is more (less) productive than production unit 0; if $\Pi(x^0,x^1,y^0,y^1,x,y,S^t)$ equals one, then the units have equal levels of productivity. At this level of generality, the index defined by (28) is due to Bjurek (1996; 308). Special cases of this type of index were described by Hicks (1961; 22), Moorsteen (1961; 462) and Diewert (1992; 240). The mathematical properties of Π with respect to x^0,x^1,y^0,y^1 are of course determined by the mathematical properties of the input index $Q(x^0,x^1,y,t)$ with respect to x^0 and x^1 and the mathematical properties of the output index $q(y^0,y^1,x,t)$ with respect to y^0 and y^1; see sections 2 and 3 above for these properties. It can be verified that if N = 1 and M = 1 so that there is only one input and one output, then the Bjurek productivity index collapses to $[y^1/y^0]/[x^1/x^0]$, which is also equal to $[y^1/x^1]/[y^0/x^0]$, the growth in Total Factor Productivity going from the production unit 0 inputs and outputs to the production unit 1 inputs and outputs. Thus if

$\Pi(x^0,x^1,y^0,y^1,x,y,S^t)$ is greater than one, production unit 1 can produce more aggregate output per unit aggregate input than production unit 0.

As usual when faced with a family of indexes, we need to determine which member of the family should be chosen for empirical applications. Again following the lead of CCD and Bjurek (1996; 310), it is natural to pick the two members of the family of indexes defined by (28) that are of most relevance to the two production units being compared. The most relevant productivity comparison for unit 0 is the Laspeyres version of (28), which is Π_L defined below by (29), where we pick the reference output and input vectors, y and x, to be the observed vectors for unit 0, y^0 and x^0, and we pick the reference technology set S^t to be S^0, the best practice technology set for production unit 0. Thus define the *Bjurek-Laspeyres productivity index* between production units 0 and 1 as:

$$(29)\ \Pi_L(x^0,x^1,y^0,y^1) \equiv q(y^0,y^1,x^0,S^0)/Q(x^0,x^1,y^0,S^0)$$
$$= [d^0(y^1,x^0)/d^0(y^0,x^0)]/[D^0(y^0,x^1)/D^0(y^0,x^0)].$$

Similarly, the most relevant productivity comparison for unit 1 is the Paasche version of (28), Π_P defined below by (30), where we pick the reference output and input vectors, y and x, to be the observed vectors for unit 1, y^1 and x^1, and we pick the reference technology set S^t to be S^1, the best practice technology set for production unit 1. Thus define the *Bjurek-Paasche productivity index* between production units 0 and 1 as:

$$(30)\ \Pi_P(x^0,x^1,y^0,y^1) \equiv q(y^0,y^1,x^1,S^1)/Q(x^0,x^1,y^1,S^1)$$
$$= [d^1(y^1,x^1)/d^1(y^0,x^1)]/[D^1(y^1,x^1)/D^1(y^1,x^0)].$$

Finally, Bjurek (1996; 310-311) suggested that a good productivity index would result if we took the geometric mean of the indexes defined by (29) and (30). Thus we define the *Bjurek productivity index* as follows:

$$(31)\ \Pi_B(x^0,x^1,y^0,y^1) \equiv [\Pi_L(x^0,x^1,y^0,y^1)\Pi_P(x^0,x^1,y^0,y^1)]^{1/2}$$
$$= \{[d^0(y^1,x^0)/d^0(y^0,x^0)][D^0(y^0,x^0)/D^0(y^0,x^1)][d^1(y^1,x^1)/d^1(y^0,x^1)][D^1(y^1,x^0)/D^1(y^1,x^1)]\}^{1/2}.$$

Thus the Bjurek productivity index is the product of two sets of output distance function ratios times two sets of input distance function ratios—a rather complicated function.

When comparing the productivity levels of two production units, it is very useful to have the productivity measure satisfy the time reversal property; i.e., if we have a productivity measure $\Pi(x^0,x^1,y^0,y^1)$ that compares the productivity level of production unit 1, characterized by the input-output data (x^1,y^1), with the productivity level of production unit 0, characterized by the input-output data (x^1,y^1), then the comparison should not depend materially on which unit is being compared to which; i.e., it would be desirable if the productivity measure satisfied the following *time reversal property*:

$$(32)\ \Pi(x^1,x^0,y^1,y^0) = 1/\Pi(x^0,x^1,y^0,y^1).$$

It is straightforward to establish the following counterpart to propositions 1 and 3 above:

Proposition 5: The Bjurek productivity index $\Pi_B(x^0,x^1,y^0,y^1)$ defined by (31) above is the *only* productivity index satisfying the time reversal property (32) that is a homogeneous symmetric average of the Laspeyres and Paasche productivity indexes, $\Pi_L(x^0,x^1,y^0,y^1)$ and $\Pi_P(x^0,x^1,y^0,y^1)$ defined by (29) and (30).

Having established its mathematical properties, our goal is now to decompose (31) into the product of readily interpreted explanatory factors; namely changes in the technical efficiency of the production units, technical progress due to a change in the reference best practice technology set from S^0 to S^1 and a measure of returns to scale. Hence in the following sections, we will consider some definitions for these explanatory variables based on distance function representations.

5. Radial Measures of Technical Efficiency

Our measures of technical efficiency for the two production units being compared are variants of the conventional Debreu (1951) Farrell (1957; 254) radial measure of efficiency loss except that in what follows we use output measures of loss rather than the input oriented measures they used.

We suppose that there are best practice technology sets S^0 and S^1 (satisfying properties P1 and P5–P7) that are relevant for production units 0 and 1. We assume that $y^t > 0_M$ and $x^t >> 0_N$ for t = 0,1 and that:

$$(33)\ (y^0,x^0)\in S^0\ ;\ (y^1,x^1)\in S^1.$$

For production units t = 0,1, the *output technical efficiency* of unit t, ε^t, is defined as follows:[22]

$$(34)\ \varepsilon^t \equiv d^t(y^t,x^t) \equiv \min_\delta \{\delta: (y^t/\delta,x^t)\in S^t\} \leq 1$$

where the inequalities in (34) follow from assumptions (33) using a feasibility argument.[23]

If $\varepsilon^0 = 1$, then production unit 0 is regarded as being efficient since the point (y^0,x^0) is on the frontier of the period 0 best practice production possibilities set. Similarly, if $\varepsilon^1 =1$, then production unit 1 is regarded as being efficient. Alternatively, if $\varepsilon^0 < 1$, then production unit 0 is clearly not efficient since an efficient period 0 producer could produce the output vector y^0/ε^0 which is strictly greater than y^0 for all positive

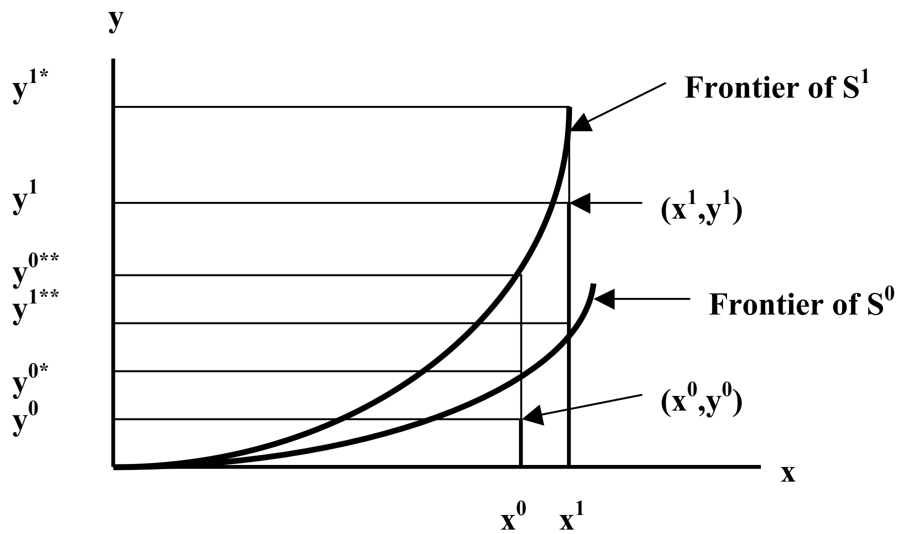
[22] This measure of technical efficiency was used by Färe, Grosskopf and Lovell (1994) and Balk (2001; 163).

[23] Our regularity conditions also imply that $\varepsilon^0 > 0$ and $\varepsilon^1 > 0$.

components of y^0, using the same input vector x^0. The amount that ε^0 is less than one is a quantitative indicator of the inefficiency of production unit 0.[24]

For the case of a single output and a single input, the technical efficiency measures can be illustrated in Figure 3.

Figure 3: Decomposition Factors for the One Output One Input Case



The observed input for unit 0 is x^0 and the corresponding amount of output produced is y^0. Note that this point lies below the frontier of the period 0 best practice technology set, S^0. The best practice technology can produce $y^{0*} > y^0$ units of output, using the same amount of input x^0. Thus the technical efficiency of production unit 0 is $\varepsilon^0 \equiv d^0(y^0,x^0) \equiv \min_\delta \{\delta : (y^0/\delta,x^0)\in S^0\} = \delta^0 = y^0/y^{0*} < 1$. Similarly, the observed input for production unit 1 is x^1 and the corresponding amount of output produced is y^1. This point lies below the frontier of the period 1 best practice technology set, S^1. The best practice technology can produce $y^{1*} > y^1$ units of output, using the same amount of input x^1. Thus the technical efficiency of production unit 1 is $\varepsilon^1 \equiv d^1(y^1,x^1) \equiv \min_\delta \{\delta: (y^1/\delta,x^1)\in S^1\} = \delta^1 = y^1/y^{1*} < 1$.

We now turn our attention to defining measures of technical change.

[24] The problem with this radial measure of inefficiency is that we could have ε^0 or ε^1 equal to one (indicating that production unit 0 or 1 is efficient) but in fact, production need not be completely efficient. This problem can be illustrated using Figure 2 where it can be seen that y^1/δ^1 is on the frontier of the reference production possibilities set S^t but it is clear that y^1/δ^1 is not completely efficient since we could use the same reference input vector to produce a greater amount of output 1 without reducing the production of output 2. This problem and possible solutions are discussed in depth by Russell and Schworm (2009) (2011). In the present paper, we will work with the rather weak measures of technical efficiency defined by (34) for the sake of simplicity but this limitation of our analysis should be kept in mind.

6. The Measurement of Technical Change

In this section, we want to use output distance functions in order to construct measures indicating by how much the reference technology changes going from period 0 to 1.[25]

Let the reference technology sets S^0 and S^1 satisfy properties P1 and P5–P7. Assume that the reference input vector x is strictly positive and that the reference output vector is nonnegative and nonzero. Then the two output distance functions $d^0(y,x)$ and $d^1(y,x)$ are well defined by (2) above and we can use these functions to define the following family of *Malmquist output based technical change measures*:[26]

(35) $\tau(y,x,S^0,S^1) \equiv d^0(y,x)/d^1(y,x)$.

Recall that the set of outputs y that are producible by the input vector x using the period t technology set S^t was denoted by $S^t(x) \equiv \{y: (y,x)\in S^t\}$ for t = 0,1. It turns out that $\tau(y,x,S^0,S^1)$ defined by (35) is a radial measure of how much bigger (or smaller) the set $S^1(x)$ is relative to $S^0(x)$; i.e., if $\tau(y,x,S^0,S^1) > 1$, then $S^0(x)$ is a strict subset of $S^1(x)$ and if if $\tau(y,x,S^0,S^1) < 1$, then $S^1(x)$ is a strict subset of $S^0(x)$ as the proof of the following proposition will show. Thus if $\tau(y,x,S^0,S^1)$ is greater (less) than one, then we have technological progress (regress) in the best practice technology going from period 0 to 1. Note also that if we reverse the role of time, then we obtain the reciprocal of the original measure of technical change; i.e., $\tau(y,x,S^1,S^0) = 1/\tau(y,x,S^0,S^1)$.

Proposition 6: Let $x \gg 0_N$ and $y > 0_M$ and suppose that the reference technology sets S^0 and S^1 satisfy properties P1 and P5–P7. Suppose that $S^0(x)$ is a subset of $S^1(x)$ so that the best practice technology does not suffer from technical regress at the reference input vector x. Then $\tau(y,x,S^0,S^1) \geq 1$. Conversely, suppose that $S^1(x)$ is a subset of $S^0(x)$. Then $\tau(y,x,S^0,S^1) \leq 1$.

Proof: Let $x \gg 0_N$ and $y > 0_M$ and suppose that $S^0(x) \subset S^1(x)$. Using definition (35), we have:

$$
\begin{aligned}
(36)\ \tau(y,x,S^0,S^1) &\equiv d^0(y,x)/d^1(y,x) \\
&= \min_\delta \{\delta: (y/\delta,x)\in S^0\}/\min_\delta \{\delta: (y/\delta,x)\in S^1\} \\
&= \delta^0/\delta^1
\end{aligned}
$$

where $(y/\delta^0,x)\in S^0$ and $(y/\delta^1,x)\in S^1$ and $\delta^0 > 0$, $\delta^1 > 0$. Note that $y/\delta^0\in S^0(x)$ and since $S^0(x) \subset S^1(x)$, it can be seen that $y/\delta^0\in S^1(x)$ and hence, δ^0 is feasible for the minimization problem $\min_\delta \{\delta: (y/\delta,x)\in S^1\} = \delta^1$. Thus $0 < \delta^1 \leq \delta^0$ and $\tau(y,x,S^0,S^1) \geq 1$. The second half of the Proposition follows in an analogous manner. Q.E.D.

[25] In the context of cross sectional comparisons of efficiency, we want to compare the best practice technology set in e.g. region 0 with the corresponding best practice set in region 1.
[26] This definition was used by Färe, Grosskopf and Lovell (1994) and Balk (2001; 163).

The technical progress measure defined by (35) is not completely satisfactory because it is local in nature. The problem is that there could be an outward shift in the reference production possibilities set going from period 0 to 1 that occurs in parts of the best practice technology sets but not at the particular (y,x) point that appears in our definition of $\tau(y,x,S^0,S^1)$. However, this problem will be minimized by careful selection of the reference (y,x) as will be seen in the following paragraph.

It is useful to choose particular cases of the general measures of technical progress defined by (35) that are most relevant to the production units being compared. Thus it is natural to choose as the reference output and input vectors, y and x, the observed output and input vectors for production units 0 and 1. This leads to the following *Laspeyres and Paasche type measures of technical progress*:[27]

(37) $\tau_L \equiv \tau(y^0,x^0,S^0,S^1) \equiv d^0(y^0,x^0)/d^1(y^0,x^0)$;
(38) $\tau_P \equiv \tau(y^1,x^1,S^0,S^1) \equiv d^0(y^1,x^1)/d^1(y^1,x^1)$.

These measures can be illustrated in the one output, one input case using Figure 3 above.

We start by analyzing the Laspeyres type measure of technical progress defined by (37) above. Note that (x^0,y^0) lies below the period 0 best practice frontier. We need to hold x^0 constant and increase y^0 to y^{0*} so that the resulting input and output combination, (x^0,y^{0*}) lies on the period 0 best practice frontier. The distance $d^0(y^0,x^0) \equiv \delta^{0*}$ will deflate y^0 onto the period 0 frontier; i.e., we have $y^0/\delta^{0*} = y^{0*}$ so that $\delta^{0*} = y^0/y^{0*}$. Next we need to hold x^0 constant and increase y^0 to y^{0**} so that the resulting input and output combination, (x^0,y^{0**}) lies on the period 1 best practice frontier. The distance $d^1(y^0,x^0) \equiv \delta^{0**}$ will deflate y^0 onto the period 1 frontier; i.e., we have $y^0/\delta^{0**} = y^{0**}$ so that $\delta^{0**} = y^0/y^{0**}$. Thus we have $\tau_L = d^0(y^0,x^0)/d^1(y^0,x^0) = \delta^{0*}/\delta^{0**} = [y^0/y^{0*}]/[y^0/y^{0**}] = y^{0**}/y^{0*}$ and it can be seen that this is a perfectly sensible proportional measure of the increase in output that is producible by the best practice technology going from period 0 to 1, using x^0 as the reference amount of input.

The analysis of the Paasche type measure of technical progress defined by (38) above proceeds in a similar manner. Note that (x^1,y^1) lies below the period 1 best practice frontier. We need to hold x^1 constant and increase y^1 to y^{1*} so that the resulting input and output combination, (x^1,y^{1*}) lies on the period 1 best practice frontier. The distance $d^1(y^1,x^1) \equiv \delta^{1*}$ will deflate y^1 onto the period 1 frontier; i.e., we have $y^1/\delta^{1*} = y^{1*}$ so that $\delta^{1*} = y^1/y^{1*}$. Next we need to hold x^1 constant and deflate y^1 to y^{1**} so that the resulting input and output combination, (x^1,y^{1**}) lies on the period 0 best practice frontier. The distance $d^0(y^1,x^1) \equiv \delta^{1**}$ will deflate y^1 onto the period 0 frontier; i.e., we have $y^1/\delta^{1**} = y^{1**}$ so that $\delta^{1**} = y^1/y^{1**}$. Thus we have $\tau_P = d^0(y^1,x^1)/d^1(y^1,x^1) = \delta^{1**}/\delta^{1*} = [y^1/y^{1**}]/[y^1/y^{1*}] = y^{1*}/y^{1**}$ and it can be seen that this is a reasonable proportional measure of the increase in output that is producible by the best practice technology going from period 0 to 1, using x^1 as the reference amount of input.

[27] These indexes were defined in CCD (1982; 1402) but were labelled as productivity indexes.

We would like a measure of technical change that is a symmetric average of the Laspeyres and Paasche measures, τ_L and τ_P defined above by (37) and (38). As usual, taking a geometric average often has nice properties. Thus define a *Fisher* (1922) *type measure of technical change* τ_F as the geometric average of τ_L and τ_P:

(39) $\tau_F \equiv [\tau_L\tau_P]^{1/2}$.

Note that both τ_L and τ_P satisfy a time reversal test; i.e., we have:

(40) $\tau_L = \tau(y^0,x^0,S^0,S^1) = 1/\tau(y^0,x^0,S^1,S^0)$;
(41) $\tau_P = \tau(y^1,x^1,S^0,S^1) = 1/\tau(y^1,x^1,S^1,S^0)$.

It can be seen that there is a counterpart to propositions 1, 3 and 5 above in the present context: the only measure of technical change that is a homogeneous symmetric average of τ_L and τ_P and also satisfies the time reversal test is the Fisher measure of technical change τ_F defined by (39).

Our final factor for the explanation of productivity change between two production units is returns to scale and we now turn to a discussion of possible measures of returns to scale.

7. Global Measures of Returns to Scale

The period 0 best practice technology will exhibit increasing returns to scale if increases in the rate of growth of inputs lead to a proportionally greater rate of growth in outputs for input-output combinations on the frontier of S^0. This concept can be illustrated in the case of one output and one input by using Figure 3. All of the Malmquist input indexes in the case of one input will be equal to x^1/x^0. For our measure of output growth, we cannot use the observed output growth ratio y^1/y^0 because the points (x^0,y^0) and (x^1,y^1) are not on the frontier of S^0. However, the points (x^0,y^{0*}) and (x^1,y^{1**}) are on the frontier of S^0 and it can be seen that our desired measure of period 0 efficient output growth (which corresponds to the input growth rate of x^1/x^0) is y^{1**}/y^{0*}. Thus in the case of one output and one input, our *Laspeyres type measure of returns to scale*, ρ_L, is defined to be $[y^{1**}/y^{0*}]/[x^1/x^0]$. Note that y^{1**} is y^1 divided by the output distance $\delta^{1**} \equiv d^0(y^1,x^1)$ so that $(y^1/\delta^{1**},x^1)$ is on the frontier of S^0. Note also that y^{0*} is y^0 divided by the output distance $\delta^{0*} \equiv d^0(y^0,x^0)$ so that $(y^0/\delta^{0*},x^0)$ is also on the frontier of S^0. We will use these output distance functions to project the observed output vectors y^0 and y^1 onto the frontier of the period 0 best practice technology in the general case of many outputs and many inputs. Thus assume S^0 satisfies P1 and P5–P7, $y^0 > 0_M$, $y^1 > 0_M$, $x^0 >> 0_N$ and $x^1 >> 0_N$. Define the *projections* of y^0 and y^1 onto the efficient period 0 best practice frontier S^0, y^{0*} and y^{1**}, as follows:

(42) $y^{0*} \equiv y^0/d^0(y^0,x^0)$; $y^{1**} \equiv y^1/d^0(y^1,x^1)$.

Our *Laspeyres type measure of returns to scale*, ρ_L, is defined to be the Laspeyres type Malmquist output index, $q_L(y^{0*},y^{1**})$, defined by (24) above, divided by the Laspeyres type Malmquist input index, $Q_L(x^0,x^1)$, defined by (6) above:

$$(43)\ \rho_L \equiv [d^0(y^{1**},x^0)/d^0(y^{0*},x^0)]/[D^0(y^0,x^1)/D^0(y^0,x^0)]$$
$$= [d^0(y^0,x^0)/d^0(y^1,x^1)][d^0(y^1,x^0)/d^0(y^0,x^0)]/[D^0(y^0,x^1)/D^0(y^0,x^0)]$$
using (42) and the linear homogeneity of $d^0(y,x^0)$ in y
$$= [\varepsilon^0/\varepsilon^1]\tau_P^{-1}\Pi_L(x^0,x^1,y^0,y^1)$$

where ε^0 and ε^1 are the technical efficiency measures for the production unit for periods 0 and 1 defined by (34), τ_P is the Paasche type measure of technical progress defined by (38) and $\Pi_L(x^0,x^1,y^0,y^1)$ is the *Bjurek-Laspeyres productivity index* between production units 0 and 1 defined by (29) above.[28] Note that if production is technically efficient in both periods so that $\varepsilon^0 = \varepsilon^1 = 1$ and if there is no technical progress between the two periods so that $\tau_P = 1$, then our returns to scale measure ρ_L collapses down to the Bjurek-Laspeyres productivity index Π_L. This is an intuitively satisfactory result.[29]

Naturally, there is a companion Paasche type measure of returns to scale that is determined by the period 1 best practice technology set S^1. From Figure 3, in this case our desired measure of period 1 efficient output growth (which corresponds to the input growth rate of x^1/x^0) is y^{1*}/ y^{0**}. Thus with one output and one input, our *Paasche type measure of returns to scale*, ρ_P, is defined to be $[y^{1*}/y^{0**}]/[x^1/x^0]$. Note that y^{1*} is y^1 divided by the output distance $\delta^{1*} \equiv d^1(y^1,x^1)$ so that $(y^1/\delta^{1*},x^1)$ is on the frontier of S^1. Note also that y^{0**} is y^0 divided by the output distance $\delta^{0**} \equiv d^1(y^0,x^0)$ so that $(y^0/\delta^{0**},x^0)$ is also on the frontier of S^1. We will use these output distance functions to project the observed output vectors y^0 and y^1 onto the frontier of the period 1 best practice technology in the general case of many outputs and many inputs. Thus assume S^1 satisfies P1 and P5–P7, $y^0 > 0_M$, $y^1 > 0_M$, $x^0 >> 0_N$ and $x^1 >> 0_N$. Define the *projections* of y^0 and y^1 onto the efficient period 1 best practice frontier S^1, y^{0**} and y^{1*}, as follows:

$$(44)\ y^{0**} \equiv y^0/d^1(y^0,x^0)\ ;\ y^{1*} \equiv y^1/d^1(y^1,x^1).$$

Our *Paasche type measure of returns to scale*, ρ_P, is defined to be the Paasche type Malmquist output index, $q_P(y^{0*},y^{1**})$, defined by (25) above, divided by the Paasche type Malmquist input index, $Q_P(x^0,x^1)$, defined by (7) above:

$$(45)\ \rho_P \equiv [d^1(y^{1*},x^1)/d^1(y^{0**},x^1)]/[D^1(y^1,x^1)/D^1(y^1,x^0)]$$
$$= [d^1(y^0,x^0)/d^1(y^1,x^1)][d^1(y^1,x^1)/d^1(y^0,x^1)]/[D^1(y^1,x^1)/D^1(y^1,x^0)]$$
using (44) and the linear homogeneity of $d^1(y,x^1)$ in y
$$= [\varepsilon^0/\varepsilon^1]\tau_L^{-1}\Pi_P(x^0,x^1,y^0,y^1)$$

[28] Let N = 1 and M = 1. Then using the linear homogeneity properties of $d^0(y,x^0)$ in y and the linear homogeneity properties of $D^0(y^0,x)$ in x, it can be seen from the first equation in (43) that $\rho_L = [y^{1**}/y^{0*}]/[x^1/x^0]$; see Figure 3 for the geometric interpretation of this measure of returns to scale.

[29] This result is consistent with Balk's (2001; 160) intuition as well, as the following quotation indicates. "Suppose that the technology, that is the feasible set of input-output quantity combinations, does not change, and that the firm is technically efficient, that is, operates on the boundary. Then the firm's productivity, in a broad sense conceived as the 'quantity' of aggregate output divided by the 'quantity' of aggregate input, can nevertheless change by moving *along* the boundary and making use of its curvature."

where ε^0 and ε^1 are again the technical efficiency measures for the production unit, τ_L is the Laspeyres type measure of technical progress defined by (37) and $\Pi_P(x^0,x^1,y^0,y^1)$ is the *Bjurek-Paasche productivity index* between production units 0 and 1 defined by (30) above. Note that if production is technically efficient in both periods so that $\varepsilon^0 = \varepsilon^1 = 1$ and if there is no technical progress between the two periods so that $\tau_L = 1$, then our returns to scale measure ρ_P collapses down to the Bjurek-Paasche productivity index Π_P.

Note that our measures of returns to scale have a "global" nature to them; i.e., they look at the average rate of growth of aggregate output between the two production units divided by the corresponding rate of growth of aggregate input where the output vectors are scaled to be on the efficient frontiers (on the frontier of S^0 for the Laspeyres measure and on the frontier of S^1 for the Paasche measure). These measures of returns to scale are different from the local measures of returns to scale introduced by CCD, which relied on differentiability of the production surfaces. In our present approach, it is not necessary to make any differentiability assumptions.

Suppose that the period 0 best practice technology S^0 is a cone (in addition to satisfying the regularity conditions P1–P7).[30] If $x^1 = \alpha x^0$ and $y^1 = \beta y^0$ where α and β are positive scalars (so that inputs and outputs grow in a proportional manner between the two observations but at possibly different rates), then our Laspeyres type measure of returns to scale should equal one under these conditions. Similarly, if the period 1 best practice technology S^1 is a cone, then our Paasche type measure of returns to scale should also equal one if outputs and inputs grow in a proportional manner. It turns out that our measures of returns to scale have these desirable properties as the following Proposition shows.

Proposition 7: Suppose the period 0 and period 1 best practice production possibility sets S^0 and S^1 satisfy the regularity conditions P1–P7 and in addition, S^0 and S^1 are cones. Let $x^0 \gg 0_N$ and $y^0 > 0_M$. Suppose in addition, that the two production units being compared have *proportional output and input vectors*; i.e., there exist $\alpha > 0$ and $\beta > 0$ such that

(46) $x^1 = \alpha x^0$; $y^1 = \beta y^0$.

Then the Laspeyres and Paasche type measures of returns to scale defined by (43) and (45) are equal to one; i.e., we have $\rho_L = 1$ and $\rho_P = 1$.

Proof: From the Appendix, we know that $d^0(y,x)$ is linearly homogeneous in the components of y and homogeneous of degree minus one in the components of x, where the latter property follows using the assumption that S^0 is a cone. We also know that $D^0(y,x)$ is linearly homogeneous in the components of x. We shall use these homogeneity properties in the proof below. Using definition (43), we have:

(47) $$\rho_L \equiv [d^0(y^0,x^0)/d^0(y^1,x^1)][d^0(y^1,x^0)/d^0(y^0,x^0)]/[D^0(y^0,x^1)/D^0(y^0,x^0)]$$
$$= [d^0(y^0,x^0)/d^0(\beta y^0,\alpha x^0)][d^0(\beta y^0,x^0)/d^0(y^0,x^0)]/[D^0(y^0,\alpha x^0)/D^0(y^0,x^0)] \quad \text{using (46)}$$

[30] S^0 is a cone if and only if $(y,x)\in S^0$ and $\lambda > 0$ implies $(\lambda y,\lambda x)\in S^0$.

$= [d^0(y^0,x^0)/\beta\alpha^{-1}d^0(y^0,x^0)][\beta d^0(y^0,x^0)/d^0(y^0,x^0)]/[\alpha D^0(y^0,x^0)/D^0(y^0,x^0)]$
$= [\alpha/\beta][\beta]/[\alpha]$
$= 1.$

The proof that $\rho_P = 1$ is similar. Q.E.D.

The above Proposition shows that our definitions of the two returns to scale measures are sensible.

It should be noted that many authors further decompose their measures of returns to scale into the product of a *pure measure of returns to scale* (where outputs and inputs grow proportionally over the two periods being compared) times *output and input mix effects*.[31] We will explain why it is not necessary to do this decomposition in our framework.

Recall our definition of the Laspeyres measure of returns to scale, ρ_L defined by (43). The numerator of this measure was the output growth index, $d^0(y^{1**},x^0)/d^0(y^{0*},x^0)$, and the denominator of this measure was the input growth measure, $D^0(y^0,x^1)/D^0(y^0,x^0)$. In order to define a "pure" measure of returns to scale, the output vector y^{1**} in the numerator of the output growth measure should be replaced by an "equivalent" output vector that is proportional to the denominator output vector y^{0*}. Similarly, the input vector x^1 in the numerator of the input growth measure should be replaced by an "equivalent" input vector that is proportional to the denominator input vector x^0. These equivalent output and input vectors, β^1y^0 and α^1x^0, are defined by the following two equations:

(48) $D^0(y^0,\alpha^1x^0) = D^0(y^0,x^1)$;
(49) $d^0(\beta^1y^{0*},x^0) = d^0(y^{1**},x^0)$.

Using the homogeneity properties of $d^0(y,x)$ in y and of $D^0(y,x)$ in x, the scalars α^1 and β^1 are defined as follows:

(50) $\alpha^1 \equiv D^0(y^0,x^1)/D^0(y^0,x^0)$; $\beta^1 \equiv d^0(y^{1**},x^0)/d^0(y^{0*},x^0)$.

Define the *pure Laspeyres measure of returns to scale*, ρ_L^*, as follows:

(51) $\rho_L^* \equiv [d^0(\beta^1y^{0*},x^0)/d^0(y^{0*},x^0)]/[D^0(y^0,\alpha^1x^0)/D^0(y^0,x^0)] = \beta^1/\alpha^1$.

Finally, *Laspeyres measures of output and input mix change*, θ_L and η_L, can be defined as follows:

(52) $\theta_L \equiv d^0(y^{1**},x^0)/d^0(\beta^1y^{0*},x^0)$; $\eta_L \equiv D^0(y^0,x^1)/D^0(y^0,\alpha^1x^0)$.

These mix effects show the effects on the output and input levels for period 1 of the switch from the output vector β^1y^{0*} to the equivalent output vector y^{1**} and of the switch from the input vector α^1x^0 to the equivalent input vector x^1.

[31] See for example Balk (2001; 170), Nemoto and Goto (2005; 620) and O'Donnell (2010; 533) (2012; 263).

Definitions (43), (51), (52) and straightforward substitutions show that our initial definition of the Laspeyres measure of returns to scale, ρ_L, has the following decomposition:

(53) $\rho_L = [\theta_L/\eta_L]\rho_L^*$.

Thus our "impure" measure of returns to scale can be decomposed into the product of mix effects times a "pure" measure of returns to scale. However, definitions (52) and the equalities (48) and (49) show that $\theta_L = \eta_L = 1$ and hence $\rho_L = \rho_L^*$.[32] Thus when we use distance functions to define output and input aggregates, mix effects are not relevant.[33]

What happens to our measures of returns to scale if we compare unit 0 to unit 1 instead of comparing unit 1 to unit 0? Denote our original Laspeyres measure of returns to scale ρ_L as $\rho_L(1/0)$ and our original Paasche measure of returns to scale ρ_P as $\rho_P(1/0)$. Now reverse the role of time and interchange the data of the two units and interchange the reference best practice technology sets S^0 and S^1. Denote the resulting Laspeyres and Paasche type measures of returns to scale by $\rho_L(0/1)$ and $\rho_P(0/1)$. It can be shown that these new measures of returns to scale are related to the old measures in the following way:[34]

(54) $\rho_L(0/1) \equiv [d^1(y^1,x^1)/d^1(y^0,x^0)][d^1(y^0,x^1)/d^1(y^1,x^1)]/[D^1(y^1,x^0)/D^1(y^1,x^1)] = 1/\rho_P(1/0)$;
(55) $\rho_P(0/1) \equiv [d^0(y^1,x^1)/d^0(y^0,x^0)][d^0(y^0,x^0)/d^0(y^1,x^0)]/[D^0(y^0,x^0)/D^0(y^0,x^1)] = 1/\rho_L(1/0)$.

Thus when we reverse the basis for comparing the two production units, the new Laspeyres type measure of returns to scale is equal to the reciprocal of the old Paasche type measure and the new Paasche type measure is equal to the reciprocal of the old Laspeyres type measure. The relations (54) and (55) suggest (as usual) that if we want a single measure of returns to scale that is a symmetric, homogeneous mean of ρ_L and ρ_P that is invariant to the way we compare the two production units, then taking the geometric mean of ρ_L and ρ_P leads to a "best" measure of returns to scale in the present context. Thus we define a *Fisher type measure of best practice returns to scale* ρ_F as the geometric average of ρ_L and ρ_P:

(56) $\rho_F \equiv [\rho_L\rho_P]^{1/2}$.

8. The Decomposition of Malmquist Productivity Indexes into Explanatory Factors

In this section, we assume that the best practice production possibilities sets S^0 and S^1 satisfy the regularity conditions P1–P7. Our goal is to compare the productivity of two

[32] A similar result can be obtained for our "impure" Paasche measure of returns to scale.
[33] This result is due to O'Donnell (2012; 263). He also shows that mix effects are present when other methods of output and input aggregation are used. However, for our present application, we can ignore mix effects.
[34] For the new measures, use the old definitions but interchange 0 and 1 everywhere in the old definitions.

production units where the observed input vector for unit t is $x^t >> 0_N$ and the observed output vector for unit t is $y^t > 0_M$ for t = 0,1.

Recall that the Bjurek-Laspeyres productivity index between units 0 and 1 was $\Pi_L(x^0,x^1,y^0,y^1)$ defined by (29) above. Equation (43) in the previous section can be manipulated to give us the following exact expression for this productivity index:

$$(57)\ \Pi_L(x^0,x^1,y^0,y^1) = [\varepsilon^1/\varepsilon^0]\ \tau_P\ \rho_L$$

where the unit t technical efficiency measures ε^t are defined by (34), the Paasche measure of technical progress τ_P is defined by (38) and the Laspeyres measure of returns to scale ρ_L is defined by the first equation in (43). Thus we have an *exact decomposition* of the Bjurek-Laspeyres productivity measure between units 0 and 1 into the product of the relative efficiency ratio $\varepsilon^1/\varepsilon^0$ times the Paasche measure of technical change between the two best practice technologies τ_P times the Laspeyres measure of returns to scale for the period 0 best practice technology ρ_L.

In a similar fashion, recall that the Bjurek-Paasche productivity index between units 0 and 1 was $\Pi_P(x^0,x^1,y^0,y^1)$ defined by (30) above. Equation (45) in the previous section can be reorganised to give us the following exact expression for this productivity index:

$$(58)\ \Pi_P(x^0,x^1,y^0,y^1) = [\varepsilon^1/\varepsilon^0]\ \tau_L\ \rho_P$$

where the unit t technical efficiency measures ε^t are defined by (34), the Laspeyres measure of technical progress τ_L is defined by (37) and the Paasche measure of returns to scale ρ_P is defined by the first equation in (45). Thus we have an *exact decomposition* of the Bjurek-Paasche productivity measure between units 0 and 1 into the product of the relative efficiency ratio $\varepsilon^1/\varepsilon^0$ times the Laspeyres measure of technical change between the two best practice technologies τ_L times the Paasche measure of returns to scale for the period 1 best practice technology ρ_P.[35]

Recall that Bjurek's recommended productivity index, $\Pi_B(x^0,x^1,y^0,y^1)$ defined by (31), was the geometric mean of the above two productivity indexes. Using (57) and (58), we have the following exact decomposition of the Bjurek productivity index:

$$\begin{aligned}(59)\ \Pi_B(x^0,x^1,y^0,y^1) &\equiv [\Pi_L(x^0,x^1,y^0,y^1)\Pi_P(x^0,x^1,y^0,y^1)]^{1/2} \\ &= [\varepsilon^1/\varepsilon^0]\ \tau_F\ \rho_F\end{aligned}$$

where τ_F is the geometric mean of τ_L and τ_P and ρ_F is the geometric mean of ρ_L and ρ_P. The exact productivity decomposition given by (59) is our preferred decomposition of the Bjurek productivity index into explanatory factors using output distance functions.[36]

[35] Of course, a knowledge of the best practice technology sets S^0 and S^1 is required in order to be able to implement the decompositions (57) and (58).

[36] It should be noted that Balk (2001; 170-171) and O'Donnell (2012; 265) obtained counterparts to our decompositions (57)-(59) for their models which used assumptions about the existence of enveloping cone

Note that we can derive a similar decomposition based on input distance functions.[37] Specifically, our preferred exact decomposition of the Bjurek productivity index into explanatory factors using input distance functions is as follows:

(60) $\Pi_B(x^0,x^1,y^0,y^1) = [\xi^0/\xi^1]\tau_F^{*}\ \rho_F^{*}$.

The terms in (60) are defined by equations (61)-(64) below, assuming that the reference technology sets S^0 and S^1 satisfy properties P1–P4, with the reference input vector x strictly positive and the reference output vector nonnegative and nonzero, so that the input distance functions $D^0(y,x)$ and $D^1(y,x)$ are well defined by (1). For production units t = 0,1, the *input technical efficiency* of unit t, ξ^t, is defined as:

(61) $\xi^t \equiv D^t(y^t,x^t) = \max_\delta\{\delta: (y^t, x^t/\delta)\in S^t\} \geq 1$.

The family of *Malmquist input based technical change* measures can be defined as:

(62) $\tau^*(y,x,S^0,S^1) \equiv D^1(y,x)/D^0(y,x)$.

The *Fisher type measure of technical progress* is defined as the geometric mean of the Laspeyres (τ_L^*) and Paasche (τ_P^*) type measures of technical progress:

(63) $\tau_F^* \equiv [\tau_L^*\tau_P^*]^{1/2} = [\tau^*(y^0,x^0,S^0,S^1)\tau^*(y^1,x^1,S^0,S^1)]^{1/2}$.

Following similar steps as for the output oriented case presented in Section 7, we obtain the following *Fisher type measure of returns to scale* as the geometric mean of the Laspeyres (ρ_L^*) and Paasche (ρ_P^*) type measure of returns to scale:

(64) $\rho_F^* \equiv [\rho_L^*\rho_P^*]^{1/2} \equiv \{[\xi^1/\xi^0][\tau_P^*]^{-1}\Pi_L(x^0,x^1,y^0,y^1) \cdot [\xi^1/\xi^0][\tau_L^*]^{-1}\Pi_P(x^0,x^1,y^0,y^1)\}^{1/2}$.

Using (31) and (63), a reorganisation of equation (64) leads to our productivity decomposition in (60).

technologies (and hence, their explanatory factors are different from our factors). It should also be noted that our decomposition (59) collapses to a decomposition obtained by O'Donnell (2012; 265) for the case when the reference technologies coincide for the two periods under consideration. However, in the general case with technical change occurring over the two periods, O'Donnell's measure of technical progress is quite different from the one we use in equation (59).

[37] Deprins, Simar and Tulkens (1984; 247) make a strong case for the use of the input concept of technical efficiency in the context of regulated firms: "It was mentioned above that distances from any observed point to the efficiency boundary can be measured in other ways than in input terms, i.e., parallel to the input axis. In the case of the various plants of a monopolistic public enterprise such as the post office, however, there is a rational for sticking to this 'input' measure of efficiency: because of the 'obligation of service'—i.e., the obligation of serving whatever demand arises at the prevailing prices—each of the individual plants has no control on its output; its only possible decisions are to adjust its input requirements to the traffic."

Finally, we note that a major problem with the decompositions of productivity growth developed in this paper is that it may be difficult to determine an appropriate reference technology set for each of the two periods under consideration. Here we simply note that, based on the Free Disposal Hull methods that were pioneered by Tulkens and his co-authors,[38] Diewert and Fox (2014) developed a DEA-type approach for decomposing productivity growth for a panel of production units into explanatory factors using the same regularity conditions as employed here.

9. Conclusion

Our paper defined Malmquist-type output, input and productivity indexes covering two periods while making very weak free disposability assumptions on the reference technology sets for the two periods under consideration. We also developed the axiomatic properties of the Malmquist input and output indexes under our weak regularity conditions on the reference technologies and we provided a justification for taking the geometric mean of Laspeyres and Paasche type Malmquist output and input indexes.

Using an output orientation, our preferred decomposition of the Bjurek productivity index into explanatory factors is given by (59), while using an input orientation our preferred decomposition is given by (60). The three explanatory factors are: (i) improvements in the technical efficiency of the production unit under consideration; (ii) technical progress for the industry over the two periods under consideration and (iii) the effects of industry returns to scale on the production unit.

A significant limitation of our productivity decompositions is the restriction that the input vectors in the data set be strictly positive. These restrictions ensured that our input distance functions are well defined. However, depending on the nature of the reference technologies, it can be shown that it is not necessary that input vectors be strictly positive in order to obtain well defined input distance functions. Additional research is required in order to relax these positivity restrictions.[39] One possible approach to relaxing the strict positivity restrictions on input vectors would be to assume that the production units minimize costs or more accurately, that the units should be minimizing costs. It makes sense to ask regulated firms and production units that provide government nonmarket services to minimize input costs even though market prices for their outputs are unavailable.[40] In a cost minimization framework, input vectors would only be required to be weakly positive instead of strictly positive.

[38] See Deprins, Simar, and Tulkens (1984) and Tulkens (1993).

[39] Consider replacing axiom P2 by the following more flexible axiom: $P2^*$. For every $y > 0_M$ and $x \equiv [x_1,...,x_N] > 0_N$, there exists $x^* \equiv [x_1^*,...,x_N^*] > 0_N$ such that $(y^*,x^*) \in S$ and $x_n^* = 0$ if $x_n = 0$ for $n = 1,...,N$. It is straightforward to modify the proof of part (i) of Proposition 8 in the Appendix and show that the input distance function, $D(y,x)$ is well defined as $\max_{\delta>0} \{\delta: (y,x/\delta) \in S\}$ for $y > 0_M$ and $x > 0_N$ using $P2^*$ in place of P2. But the new axiom allows for too much flexibility in production; i.e., it rules out some inputs being *essential* for the production of a positive output vector. It seems best to leave the development of a less restrictive version of axiom P2 to an actual empirical application.

[40] Diewert (2011) (2012) developed a cost based approach to productivity measurement in nonmarket contexts but instead of using a nonparametric approach with respect to outputs, he used an index number

Appendix: **Regularity Conditions on the Reference Technology and Properties of Distance Functions**

In order to simplify the notation, we will drop the superscript t in what follows. We assume that the production possibilities set S is given and for $y > 0_M$ and $x >> 0_N$, the *input distance function* D and the *output distance function* d are defined as follows:

(A1) $D(y,x) \equiv \max_{\delta>0} \{\delta: (y,x/\delta)\in S\}$.
(A2) $d(y,x) \equiv \min_{\delta>0} \{\delta: (y/\delta,x)\in S\}$.

Diewert and Fox (2010) showed that if $y > 0_M$ and $x >> 0_N$ and S satisfies properties P1–P4, then the input distance function D(y,x) is well defined as the maximum in (A1) with $D(y,x) > 0$. They also showed that if $y >> 0_M$ and $x >> 0_N$ and S satisfies properties P1 and P5–P7, then the output distance function d(y,x) is well defined as the minimum in (A2) with $d(y,x) > 0$. Note that these results did not require any convexity assumptions on the technology set S.

Variants of the following propositions are well known in the literature but our regularity conditions are a bit weaker than the convexity conditions used by others.[41] Note that the restriction used by Diewert and Fox (2010) that $y >> 0_M$ is relaxed to the weaker restriction that $y > 0_M$ in Proposition 9 below.

Proposition 8: Suppose the production possibilities set S satisfies properties P1–P4 listed above. Suppose the reference output vector y satisfies $y > 0_M$ and define the positive orthant in N dimensional Euclidean space by $\Omega_N \equiv \{x : x >> 0_N\}$. Then the input distance function D(y,x) defined by (A1) above is (i) well defined and positive, (ii) nondecreasing, (iii) positively linearly homogeneous, (iv) increasing if all inputs increase and (v) continuous in x over Ω_N.

Proof of (i): Follows from Proposition 2 in Diewert and Fox (2010).[42]

Proof of (ii): Let $y > 0_M$ and $0_N << x^0 < x^1$. Using definition (A1) and property (i), we have the existence of a positive scalar δ^0 such that:

(A3) $D(y,x^0) \equiv \max_{\delta>0} \{\delta: (y,x^0/\delta)\in S\} = \delta^0 > 0$

approach, where "prices" for nonmarket outputs were based on unit costs or marginal costs for producing these outputs.

[41] For a fairly comprehensive discussion of regularity conditions on the technology and the resulting properties of the input and output distance functions, see Färe (1988).

[42] There is a typographical error in their proof. A corrected proof proceeds as follows. Let $y > 0_M$ and $x >> 0_N$. Then by P2, there exists $x^* \geq 0_N$ such that $(y,x^*)\in S$. Since $x >> 0_N$, there exists a $\delta^* > 0$ that is small enough such that $x/\delta^* \geq x^*$. Thus by P3, $(y,x/\delta^*)\in S$. We cannot increase δ^* to plus infinity and conclude that $(y,0_N)\in S$ because this would contradict P4. Using the fact that S is a closed set, it can be seen that the maximization problem defined by (A1) has a finite positive maximum, δ^{**}.

where $(y,x^0/\delta^0)\in S$. Since $x^1 > x^0$ and $\delta^0 > 0$,

(A4) $x^1/\delta^0 > x^0/\delta^0$.

Since $(y,x^0/\delta^0)\in S$ and (A4) holds, property P3 implies that

(A5) $(y,x^1/\delta^0)\in S$.

Thus

(A6) $D(y,x^1) \equiv \max_{\delta>0} \{\delta: (y,x^1/\delta)\in S\}$
$\geq \delta^0$ since by (A5) δ^0 is feasible for the maximization problem
$= D(y,x^0)$ using (A3).

Proof of (iii): Let $y > 0_M$, $x >> 0_N$ and $\lambda > 0$. Then using property (i) and definition (A1), we have the existence of a positive scalar δ^* such that

(A7) $D(y,x) \equiv \max_{\delta>0} \{\delta: (y,x/\delta)\in S\} = \delta^* > 0$.

Thus $(y,x/\delta^*)\in S$ and since $\lambda > 0$, we also have

(A8) $(y,\lambda x/\lambda\delta^*)\in S$.

Now calculate the value of the input distance function $D(y,\lambda x)$:

(A9) $D(y,\lambda x) \equiv \max_{\varepsilon>0} \{\varepsilon: (y,\lambda x/\varepsilon)\in S\}= \varepsilon^* \equiv \lambda\delta^{**} \geq \lambda\delta^* = \lambda D(y,x)$

where the inequality follows from the feasibility of $\lambda\delta^*$ for the maximization problem in (A9); see (A8). Note that we have defined $\delta^{**} \equiv \varepsilon^*/\lambda$.

Suppose the strict inequality in (A9) holds. Then δ^{**} is such that

(A10) $D(y,\lambda x) = \lambda\delta^{**} > \lambda\delta^*$.

The equality in (A10) implies that $(y,\lambda x/\lambda\delta^{**})\in S$. But then we also have

(A11) $(y,x/\delta^{**})\in S$.

From (A7), we have

(A12) $\delta^* = \max_{\delta>0} \{\delta : (y,x/\delta)\in S\}$
$\geq \delta^{**}$ since by (A11), δ^{**} is feasible for the maximization problem
$> \delta^*$ using $\lambda > 0$ and (A10).

But (A12) is a contradiction and thus our *supposition* is false and property (iii) follows.

Proof of (iv): Let $0_N \ll x^0 \ll x^1$. Then there exists a scalar $\lambda > 1$ such that

(A13) $\lambda x^0 \leq x^1$.

Since $\lambda > 1$, we have:

(A14) $D(y,x^0) < \lambda D(y,x^0)$ since $D(y,x^0) > 0$ and $\lambda > 1$
$= D(y,\lambda x^0)$ using the linear homogeneity of D
$\leq D(y,x^1)$ using (A13) and weak monotonicity.

Proof of (v): Let $x^0 \gg 0_N$ and choose $\alpha > 0$ small enough so that $x^0 - \alpha x^0 = (1-\alpha)x^0 \gg 0_N$. Define the *hyperblock* in N dimensional space of size α that is centered around x^0 as follows:

(A15) $H(x^0,\alpha) \equiv \{x: (1-\alpha)x^0 \leq x \leq (1+\alpha)x^0\}$.

Note that $H(x^0,\alpha)$ is a subset of the positive orthant and x^0 is in the interior of $H(x^0,\alpha)$. Using the definition of $D(y,x^0)$, there exists a $\delta^0 > 0$ such that $D(y,x^0) = \delta^0$ and $(y,x^0/\delta^0)\in S$. Using the linear homogeneity property of D(y,x) in x, we have:

(A16) $D(y,(1-\alpha)x^0) = (1-\alpha)D(y,x^0) = (1-\alpha)\delta^0$;
(A17) $D(y,(1+\alpha)x^0) = (1+\alpha)D(y,x^0) = (1+\alpha)\delta^0$.

Using the weak monotonicity property of D(y,x) in x and (A16) and (A17), it can be seen that for all $x\in H(x^0,\alpha)$, we have:

(A18) $(1-\alpha)\delta^0 = D(y,(1-\alpha)x^0) \leq D(y,x) \leq D(y,(1+\alpha)x^0) = (1+\alpha)\delta^0$.

The inequalities in (A18) are sufficient to imply the continuity of D(y,x) at the point x^0.
Q.E.D.

We turn to the analysis of the properties of the output distance function, d(y,x), in y for fixed $x \gg 0_N$.

Proposition 9: Suppose the production possibilities set S satisfies properties P1 and P5–P7 listed above. Suppose the reference input vector x satisfies $x \gg 0_N$ and define the nonnegative orthant in M dimensional Euclidean space, excluding the origin, by $\Omega_M^* \equiv \{y : y > 0_M\}$. Then the output distance function d(y,x) defined by (A2) above is (i) well defined and positive for $y\in\Omega_M^*$, (ii) nondecreasing, (iii) positively linearly homogeneous, (iv) increasing if all outputs increase and (v) continuous in y over the interior of Ω_M^*; i.e., D(y,x) is continuous in y over the set of strictly positive y.

Proof of (i): Let $y > 0_M$ and $x >> 0_N$. Since $x >> 0_N$, by P6, there exists a $y^* >> 0_M$ such that $(y^*,x)\in S$. Since y^* is strictly positive and y is nonnegative but nonzero, there exists $\delta^* > 0$ large enough so that $y/\delta^* \leq y^*$. Using P7, we see that $(y/\delta^*,x)\in S$ and thus we have a feasible solution for the minimization problem in (A2). From definition (A2), we want to make $\delta \geq 0$ as small as possible such that $(y/\delta,x)\in S$. However, we cannot make $\delta > 0$ but arbitrarily close to 0 and have $(y/\delta,x)$ belong to S because this would contradict property P5. Using property P1, we see that a finite positive minimum for the minimization problem in (A2) exists.

Proof of (ii): Let $x >> 0_N$ and $0_M < y^0 \leq y^1$. Using definition (A2) and property (i), we have the existence of a positive scalar δ^1 such that:

(A13) $d(y^1,x) \equiv \min_{\delta>0} \{\delta: (y^1/\delta,x)\in S\} = \delta^1 > 0$

where $(y^1/\delta^1,x)\in S$. Since $y^0 \leq y^1$ and $\delta^1 > 0$,

(A14) $y^0/\delta^1 \leq y^1/\delta^1$.

Since $(y^1/\delta^1,x)\in S$ and (A14) holds, Property P7 implies that

(A15) $(y^0/\delta^1,x)\in S$.

Thus

(A16) $D(y^0,x) \equiv \min_{\delta>0} \{\delta: (y^0/\delta,x)\in S\}$
$\leq \delta^1$ since by (A15) δ^1 is feasible for the minimization problem
$= D(y^1,x)$ using (A13).

Proofs of (iii), (iv) and (v): Analogous to the proofs of (iii), (iv) and (v) in the previous proposition. Q.E.D.

Note that we can only establish the continuity of d(y,x) in y over the *positive orthant* in M space; our regularity conditions are not strong enough to rule out discontinuities at the boundary of the positive orthant.

The regularity conditions on S listed above do not include any convexity assumptions. If we are willing to make some convexity assumptions on the reference technology, then we can deduce some additional properties for the output and input distance functions. Thus we consider the following two additional regularity conditions on S:

P8: For each $y > 0_M$, the input possibilities set $S(y) \equiv \{x: (y,x)\in S\}$ is a convex set.
P9: For each $x >> 0_N$, the output possibilities set $S^*(x) \equiv (y: (y,x)\in S)$ is a convex set.

Note that the convexity assumptions P8 and P9 do not rule out increasing returns to scale for the reference technology S. These types of convexity assumptions are relevant if the reference technology S is generated by a DEA exercise.

Proposition 10: Suppose the technology set S satisfies properties P1–P4 and P8. Then for each $y > 0_M$, the input distance function $D(y,x)$ is a *concave function* of x over the positive orthant Ω_N.

Proof: Let $y > 0_M$. $x^1 >> 0_N$. $x^2 >> 0_N$ and $0 < \lambda < 1$. From Proposition 8, $D(y,x)$ is positive, monotonic and linearly homogeneous in x for $x \in \Omega_N$. We first show that $D(y,x)$ is a quasiconcave function of x over Ω_N. Let $D(y,x^1) = \delta^1 > 0$, $D(y,x^2) = \delta^2 > 0$ and without loss of generality, assume:

(A17) $0 < D(y,x^1) = \delta^1 \le \delta^2 = D(y,x^2)$.

Using (A17) and the linear homogeneity property of $D(y,x)$ in x, we have:

(A18) $D(y,x^1/\delta^1) = 1$; $D(y,x^2/\delta^2) = 1$

and hence $(y,x^1/\delta^1) \in S$ and $(y,x^2/\delta^2) \in S$. Hence using property P8, we have

(A19) $(y,\lambda[x^1/\delta^1] + (1-\lambda)[x^2/\delta^2]) \in S$.

Thus using definition (A1):

(A20) $D(y,\lambda[x^1/\delta^1] + (1-\lambda)[x^2/\delta^2]) \equiv \max_{\delta>0} \{\delta: (y,(\lambda[x^1/\delta^1] + (1-\lambda)[x^2/\delta^2])/\delta) \in S\} \ge 1$

since by (A19), $\delta = 1$ is feasible for the maximization problem in (A20). Thus we have

(A21) $1 \le D(y,\lambda[x^1/\delta^1] + (1-\lambda)[x^2/\delta^2])$
$\le D(y,\lambda[x^1/\delta^1] + (1-\lambda)[x^2/\delta^1])$ using (A17) and property (ii) in Proposition 8
$= [\delta^1]^{-1} D(y,\lambda x^1 + (1-\lambda)x^2)$ using property (iii) in Proposition 8.

But (A21) and (A17) shows that

(A22) $D(y,\lambda x^1 + (1-\lambda)x^2) \ge \min \{D(y,x^1), D(y,x^2)\}$

which establishes the quasiconcavity of $D(y,x)$ in x over Ω_N. We now establish the concavity of $D(y,x)$ with respect to x over Ω_N. Recall the definitions and inequalities in (A17). Define α as follows:

(A23) $\alpha \equiv D(y,x^1)/D(y,x^2) = \delta^1/\delta^2 \le 1$.

Therefore $D(y,x^1) = \alpha D(y,x^2) = D(y,\alpha x^2)$ using the linear homogeneity of $D(y,x)$ in x. Thus for all μ such that $0 \le \mu \le 1$, we have

(A24) $\min\{D(y,x^1), D(y,\alpha x^2)\} = D(y,x^1) \le D(y,\mu x^1 + (1-\mu)\alpha x^2)$

where the inequality follows using the quasiconcavity of D(y,x) in x. Now look for a $\beta > 0$ and a μ between 0 and 1 such that

(A25) $\mu x^1 + (1-\mu)\alpha x^2 = \beta[\lambda x^1 + (1-\lambda)x^2]$.

The β and μ solution to (A25) is

(A26) $\beta = \alpha/[1 - \lambda + \lambda\alpha] > 0$ and $\mu = \alpha\lambda/[1 - \lambda + \alpha\lambda]$

where μ lies between 0 and 1. Using (A24), we have

(A27)
$$\begin{aligned} D(y,x^1) &\le D(y,\mu x^1 + (1-\mu)\alpha x^2) \\ &= D(y, \beta[\lambda x^1 + (1-\lambda)x^2]) && \text{using (A25)} \\ &= \beta D(y, \lambda x^1 + (1-\lambda)x^2) && \text{using the linear homogeneity of } D(y,x) \text{ in } x. \end{aligned}$$

(A27) can be rewritten as

(A28)
$$\begin{aligned} D(y, \lambda x^1 + (1-\lambda)x^2) &\ge \beta^{-1} D(y,x^1) \\ &= [1 - \lambda + \lambda\alpha]\alpha^{-1} D(y,x^1) && \text{using (A26)} \\ &= \lambda D(y,x^1) + [1 - \lambda]\alpha^{-1} D(y,x^1) \\ &= \lambda D(y,x^1) + [1 - \lambda]D(y,x^2) && \text{using (A23)} \end{aligned}$$

which establishes the concavity of D(y,x) over Ω_N. Q.E.D.

The fact that a positive, quasiconcave and linearly homogeneous function is also concave was first established by Berge (1963).

Proposition 11: Suppose the technology set S satisfies properties P1, P5–P7 and P9. Then for each $x >> 0_N$, the output distance function d(y,x) is a *convex function* of y over the nonnegative orthant less the origin $\Omega_M{}^*$.

Proof: The proof is a straightforward modification of the proof of Proposition 10. Q.E.D.

There is one additional regularity condition that is useful to impose on the technology set S:

P10: S is a cone.[43]

[43] Thus if $(y,x)\in S$ and $\lambda \ge 0$, then $(\lambda y,\lambda x)\in S$.

Proposition 12: Suppose the technology set S satisfies properties P1–P7 and P10 and let $\lambda > 0$, $y > 0_M$ and $x >> 0_N$. Then the output and input distance functions, d(y,x) and D(y,x), have the following homogeneity properties:

(A29) $d(y,\lambda x) = \lambda^{-1}d(y,x)$;
(A30) $D(\lambda y,x) = \lambda^{-1}D(y,x)$.

Proof of (A29): Let $\lambda > 0$, $y > 0_M$ and $x >> 0_N$. Then

(A31) $d(y,x) \equiv \min_{\delta>0} \{\delta : (y/\delta,x)\in S\} = \delta^* > 0$

where $(y/\delta^*,x)\in S$. Since S is a cone by assumption, we have:

(A32) $(\lambda y/\delta^*,\lambda x) = (y/\lambda^{-1}\delta^*,\lambda x)\in S$.

Using the definition of $d(y,\lambda x)$, we have:

(A33) $d(y,\lambda x) \equiv \min_{\delta>0} \{\delta : (y/\delta,\lambda x)\in S\}$
$\leq \lambda^{-1}\delta^*$

since using (A32), it can be seen that $\lambda^{-1}\delta^*$ is feasible for the minimization problem. Now *suppose* that

(A34) $d(y,\lambda x) \equiv \delta^{**} < \lambda^{-1}\delta^*$.

Thus $(y/\delta^{**},\lambda x)\in S$. Since S is a cone, $(y/\delta^{**},\lambda x)\in S$ implies $(y/\lambda\delta^{**},x)\in S$. Thus

(A35) $d(y,x) \equiv \min_{\delta>0} \{\delta : (y/\delta,x)\in S\}$
$\leq \lambda\delta^{**}$ since $(y/\lambda^{-1}\delta^{**},x)\in S$
$< \lambda\lambda^{-1}\delta^*$ using (A34)
$= d(y,x)$.

But (A35) is impossible so our *supposition* is false and hence we have (A29). The proof of (A30) follows using analogous feasibility arguments. Q.E.D.

In Diewert and Fox (2014), we showed how input and output distance functions and productivity decompositions could be computed where the reference best practice technologies were generated by the Free Disposal Hulls of a finite set of observations on production units for each of two periods.[44] It may be useful to indicate how to calculate the input and output distance functions for a reference technology that consists of the *union of the Free Disposal Hulls of a number of simple linear processes*. We will indicate

[44] Of course, our decompositions drew on the pioneering work of Deprins, Simar and Tulkens (1984), Tulkens (1993) and Tulkens and Eeckaut (1995a) (1995b).

how this can be done for K processes which produce positive amounts of all M outputs and use positive amounts of all N inputs when operated at unit scale.[45]

Suppose that the strictly positive output vector $y^k \equiv [y_1^k,...,y_M^k] \gg 0_M$ can be produced by the strictly positive input vector $x^k \equiv [x_1^k,...,x_N^k] \gg 0_N$ for k = 1,...,K. The *Free Disposal Conical Hull*, S^k, generated by (y^k,x^k) is defined as the following production possibilities set:

$$\text{(A36) } S^k \equiv \{(y,x) : 0_M \le y \le y^k\lambda \,;\, x \ge x^k\lambda \,;\, \lambda \ge 0\} \,; \qquad k = 1,...,K.$$

Now calculate the input distance function, $D^k(y,x)$ for $y \equiv [y_1,...,y_M] > 0_M$ and $x \equiv [x_1,...,x_N] \gg 0_N$ that corresponds to the reference technology S^k for k = 1,...,K:

$$\begin{aligned}
\text{(A37) } \delta^k \equiv D^k(y,x) &\equiv \max_{\delta>0}\{\delta : (y,x/\delta)\in S^k\} \\
&= \max_{\delta>0,\lambda\ge 0}\{\delta : y \le y^k\lambda \,;\, x/\delta \ge x^k\lambda\} && \text{using (A36)} \\
&= \max_{\delta>0,\lambda\ge 0}\{\delta : y \le y^k\lambda \,;\, x \ge x^k\lambda\delta\} \\
&= \max_{\delta>0}\{\delta : x \ge x^k\lambda^k\delta\} && \text{where } \lambda^k \equiv \max_{m=1,...,M}\{y_m/y_m^k\} \\
&= \max_{\delta>0}\{\delta : x_n/(x_n^k\lambda^k) \ge \delta \,;\, n = 1,...,N\} \\
&= \min_{n=1,...N}\{x_n/x_n^k\}/\max_{m=1,...,M}\{y_m/y_m^k\}.
\end{aligned}$$

Note that since $x \gg 0_N$ and $y > 0_M$, each δ^k is positive. Now define the reference technology S as the *union* of the K technology sets S^k:

$$\text{(A38) } S \equiv \cup_{k=1}^K S^k \,.$$

Let D(y,x) be the *input distance function* that corresponds to S. Thus for $y > 0_M$ and $x \gg 0_N$, we have:

$$\text{(A39) } D(y,x) \equiv \max_{\delta>0}\{\delta : (y,x/\delta)\in S\} = \max_{\delta>0}\{\delta : (y,x/\delta)\in \cup_{k=1}^K S^k\} = \max_{k=1,...,K}\{\delta^k\}$$

where δ^k is defined by (A37) for k = 1,...,K.

In a similar fashion, we can show that the *output distance function* $d^k(y,x)$ for $y > 0_M$ and $x \gg 0_N$ that is generated by the production possibilities set S^k is defined as follows for k = 1,...,K:

$$\text{(A40) } d^k(y,x) \equiv \min_{\delta>0}\{\delta : (y/\delta,x)\in S^k\} = \max_{m=1,...,M}\{y_m/y_m^k\}/\min_{n=1,...N}\{x_n/x_n^k\} \equiv \phi^k$$

where each ϕ^k is positive. The *output distance function* that corresponds to the union technology set S is defined as follows for $y > 0_M$ and $x \gg 0_N$:

$$\text{(A41) } d(y,x) \equiv \min_{\delta>0}\{\delta : (y/\delta,x)\in S\} = \min_{\delta>0}\{\delta : (y/\delta,x)\in \cup_{k=1}^K S^k\} = \min_{k=1,...,K}\{\phi^k\}$$

where ϕ^k is defined in equations (A40) for each k.

[45] Relaxing these strict positivity conditions leads to considerably more complicated formulae.

It is easy to see that D(y,x) is linearly homogeneous in x and homogeneous of degree −1 in y and d(y,x) is linearly homogeneous in y and homogeneous of degree −1 in x. It can also be seen that our positivity restrictions $y > 0_M$ and $x >> 0_N$ are required in order to ensure that both distance functions are well defined as finite numbers.

References

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

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.

Berge, C. (1963), *Topological Spaces*, New York: Macmillan.

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

Briec, W. and K. Kerstens (2011), "The Hicks-Moorsteen Productivity Index Satisfies the Determinateness Axiom", *The Manchester School* 79, 765-775.

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

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

Charnes, A., W.W. Cooper and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research* 2, 429-444.

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

Debreu, G. (1951), "The Coefficient of Resource Utilization", *Econometrica* 19, 273-292.

Deprins, D., L. Simar, and H. Tulkens (1984), "Measuring Labor Efficiency in Post Offices", pp. 243-267 in M. Marchand, P. Pestieau, and H. Tulkens (eds.), *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland.

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

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

Diewert, W.E. (1993), "Symmetric Means and Choice under Uncertainty"', pp. 355-433 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

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

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

Diewert, W.E. (2012), "The Measurement of Productivity in the Nonmarket Sector", *Journal of Productivity Analysis* 37, 217-229.

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

Diewert, W.E. and K.J. Fox (2014), "Reference Technology Sets, Free Disposal Hulls and Productivity Decompositions", *Economics Letters* 122, 238–242.

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

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

Färe, R. (1988), *Fundamentals of Production Theory*, Lecture Notes in Economics and Mathematical Systems 311, Berlin: Springer-Verlag.

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

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

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

Färe, R., S. Grosskopf and P. Roos (1996), "On Two Definitions of Productivity", *Economic Letters* 53, 269-274.

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

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

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

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

Grosskopf, S. (1986), "The Role of the Reference Technology in Measuring Productive Efficiency", *Economic Journal* 96, 499-513.

Grosskopf, S. (2003), "Some Remarks on Productivity and its Decompositions", *Journal of Productivity Analysis* 20, 459-474.

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

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

Hicks, J. (1981), "The Valuation of the Social Income III—the Cost Approach", pp. 243-268 in *Wealth and Welfare*, Cambridge, MA: Harvard University Press.

Kerstens, K. and I. Van de Woestyne (2014), "Comparing Malmquist and Hicks–Moorsteen Productivity Indices: Exploring the impact of unbalanced vs. balanced panel data," *European Journal of Operational Research* 233, 749–758.

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

Malmquist, S. (1953), "Index Numbers and Indifference Surfaces", *Trabajos de Estastistica* 4, 209-242.

Moorsteen, R.H. (1961), "On Measuring Productive Potential and Relative Efficiency", *Quarterly Journal of Economics* 75, 451-467.

Nemoto, J. and M. Goto (2005), "Productivity, Efficiency, Scale Economies and Technical Change: A new decomposition analysis of TFP applied to the Japanese prefectures", *Journal of the Japanese and International Economies* 19, 617–634.

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

O'Donnell, C.J. (2012), "Nonparametric Estimates of the Components of Productivity and Profitability Change in U.S. Agriculture", *American Journal of Agricultural Economics* 94, 875-890.

Ray, S.C. and E. Desli (1997), "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries: Comment", *American Economic Review* 87, 1033-1039.

Russell, R.R. and W. Schworm (2009), "Axiomatic Foundations of Efficiency Measurement on Data Generated Technologies", *Journal of Productivity Analysis* 31, 77-86.

Russell, R.R. and W. Schworm (2011), "Properties of Inefficiency Indexes on Input, Output Space", *Journal of Productivity Analysis* 36, 143-156.

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

Tulkens, H. (1993), "On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts, and Urban Transit", *Journal of Productivity Analysis* 4, 183–210.

Tulkens, H. and P.V. Eeckaut (1995a), "Non-Frontier Measures of Efficiency, Progress and Regress for Time Series Data", *International Journal of Production Economics* 39, 83-97.

Tulkens, H. and P.V. Eeckaut (1995b), "Nonparametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects", *European Journal of Operational Research* 80, 474-499.

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

Zelenyuk, V. (2013), "Scale Efficiency and Homotheticity: Equivalence of primal and dual measures" *Journal of Productivity Analysis*, first online (DOI 10.1007/s11123-013-0361-z).