Theoretical Productivity Indices

R. Robert Russell*
University of California, Riverside

June 2017

Abstract

This chapter, scheduled for inclusion in the forthcoming *Oxford Handbook of Productivity Analysis* (edited by E. Grifell-Tatjé, C. A. K. Lovell, and R. C. Sickles), lays out the theory of productivity measurement for multiple-input, multiple-output production technologies. The foundational concepts for this theory are the Malmquist Index, developed by D. W. Caves, L. R. Christensen, and W. E. Diewert, and the Hicks-Moorsteen Index, developed by H. Bjurek. The former is fundamentally a measure of the shift of the production frontier, while the latter is the ratio of an output index to an input index. Each uses Malmquist-Shephard distance functions to measure the (radial) changes in input and output vectors. Much of the subsequent literature extends these notions to incorporate (i) the possibility of technological inefficiency (operation below the production frontier) and (ii) decomposition into components of productivity change (like efficiency change, scale effects, and input- or output-mix effects). Other developments covered in he chapter are the non-radial (hyperbolic and directional distance) indices and dual productivity indices.

*Email: robert.russell@ucr.edu

Chapter 4

Theoretical Productivity Indices

R. Robert Russell[1]

4.1 Introduction: Solow Technical Change

This chapter focuses on *theoretical* productivity indices, defined for the purpose at hand as indices predicated on the assumption that the technology is known and non-stochastic but unspecified (*e.g.*, non-parameterized). See Chapter 2 for a thorough discussion of more expansive notions of non-stochastic productivity indices.

I begin, in this introductory section, with a preview of the chapter, motivating the contents with a retrospective look at Solow's [1957] seminal contribution to the measurement of technical change, the genesis of today's analysis of productivity comparisons (across production units as well as time).

[1]This chapter was substantially improved by a thorough perusal of a wobbly first draft by Bert Balk and Knox Lovell. It has also benefitted from comments by Chris O'Donnell, Rolf Färe, and Oleg Badunenko.

A (mostly inconsequential) generalization of Solow's set-up[2] is as follows:

$$y^t = f^t(x^t) \quad \forall\, t,$$

where, for observation t, $y^t \in \mathbf{R}_+$ is the scalar output quantity, $x^t \in \mathbf{R}_+^n$ is an input quantity vector, and $f^t : \mathbf{R}_+^n \to \mathbf{R}_+$ is the production function. The nomenclature of the chapter adheres to Solow's context in which t is a time index, but the results apply equally to cross-sectional analysis, where t is an index assigned to different economic units (*e.g.*, firms or national economies).

For expositional purposes only, Solow considered the special case of neutral technological change:

$$y^t = a^t f(x^t) \quad \forall\, t, \tag{4.1}$$

where a^t is a scalar indicator of the state of technology and $f : \mathbf{R}_+^n \to \mathbf{R}_+$ is a stationary (reference) production function. Additionally, $a^t = y^t/f(x^t)$ has come to be interpreted as *total factor productivity* at observation t, where $f(x^t)$ is the (aggregate) total factor input at observation t.

Solow's definition of technical change from a base period b to a current period c, under the assumption of neutral technological change, is given simply by the relative productivity states, a^c/a^b. The Solow productivity index under neutrality, $\Pi^{SN} : \mathbf{R}_+^{n+1} \to \mathbf{R}_+$, is then obtained by substitution from (4.1):[3]

$$\Pi^{SN}(x^b, x^c, y^b, y^c) := \frac{a^c}{a^b} = \frac{y^c/f(x^c)}{y^b/f(x^b)}. \tag{4.2}$$

The last term in (4.2) underscores the interpretation of Solow's technical change index, under the assumption of neutral technological change, as an index of total-factor productivity change.

Another evocative interpretation of Solow technical change is obtained by multiplying top and bottom of a^c/a^b by $f(x^b)$ to obtain

$$\Pi^{SN}(x^b, x^c, y^b, y^c) := \frac{a^c f(x^b)}{a^b f(x^b)} = \frac{f^c(x^b)}{y^b}. \tag{4.3}$$

[2] Allowing for more than two inputs *is* inconsequential; eschewing Solow's assumption of first-degree homogeneity (constant returns to scale) of the production function is not inconsequential but is nevertheless innocuous for most of our discussion.

[3] $A := B$ means the relation defines A, and $A =: B$ means the relation defines B.

The last term indicates that Solow technical change is given by the ratio of (a) the maximal (frontier) output that can be produced with the base-period input vector using the current-period technology to (b) the (frontier) output produced in the base period. While the last equality in (4.3) holds only under the assumption of neutral technology change, the last term is a natural definition of technological change for the general case of (possibly) non-neutral technological change.

Under neutrality, the proportional shift of the frontier is identical across input vectors; under non-neutrality, when the frontier shift is input dependent, the last term in (4.3) measures the shift in the frontier specifically at the base-period input vector. In fact, as Solow did not assume neutrality in his calculations, his index was actually *defined* by

$$\Pi^{Sb}(x^b, x^c, y^b, y^c) := \frac{f^c(x^b)}{y^b}, \tag{4.4}$$

which is input-quantity dependent and indicates how much more can be produced with the base-period input vector using the current-period technology (rather than the base-period technology).

While Solow measured technological change at the base-period input quantity (using the current-period technology), he could just as well have measured it at the current-period input vector (using the base-period technology):

$$\Pi^{Sc}(x^b, x^c, y^b, y^c) := \frac{y^c}{f^b(x^c)}. \tag{4.5}$$

The indices in (4.4) and (4.5) are identical if and only if technological change is neutral; *i.e.*, if and only if the production function is given by (4.1).

The indices (4.4) and (4.5) are special cases—owing to the restriction to a single output—of the widely employed (output-oriented) *Malmquist productivity indices* formulated by Caves, Christensen, and Diewert [1982] (CCD). Generalization of (4.4) and (4.5) to multiple outputs requires an aggregation rule to obtain a measure of the change in "output." The aggregation rule employed by CCD is the Malmquist [1953] output distance (gauge) function (independently formulated by Shephard [1953]). In this context, it serves essentially as a measure of the "radial distance" from an output vector to a future or past technology frontier in output space (*i.e.*, a production possibility frontier) for a fixed input, much as, *e.g.*, $y^c/f^b(x^c)$ is a proportional

measure of the distance from current-period output to the base-period production frontier at the current-period input vector x^c.

Measuring technological change in the region of ⟨input, output⟩ space at which the economic unit is operating has intuitive appeal, but there seems to be no compelling criterion to choose between using the current-period technology and the base-period quantities (4.4) or using the base-period technology and the current-period quantities (4.5). Diewert [1992a] therefore suggests adoption of the Fisher "ideal" index number formulation—*viz.*, taking the geometric average of the two indices. In the single-output case, the *Malmquist "ideal" output-based productivity index* is

$$\Pi^{MI}(x^b, x^c, y^b, y^c) = \left(\frac{f^c(x^b)}{y^b}\right)^{1/2} \left(\frac{y^c}{f^b(x^c)}\right)^{1/2}.$$

As the above nomenclature suggests, Diewert also formulated *input-based* Malmquist and Malmquist "ideal" indices. These formulations employ the Malmquist-Shephard input-oriented distance function, a radial measure of the distance of an input vector from a base-period or current-period frontier in input space (*i.e.*, an isoquant).

The CCD-Diewert development of the various Malmquist productivity indices in the general case of multiple outputs and multiple inputs is described in Section 4.3 (following presentation of some preliminaries in Section 4.2).

Suppose we multiply top and bottom of the indices in (4.4) and (4.5) by $y^c = f^c(x^c)$ and $y^b = f^b(x^b)$, respectively, to obtain

$$\Pi^{Sb}(x^b, x^c, y^b, y^c) = \frac{y^c/y^b}{f^c(x^c)/f^c(x^b)} \tag{4.6}$$

and

$$\Pi^{Sc}(y^c, y^b, x^c, x^b) = \frac{y^c/y^b}{f^b(x^b)/f^b(x^c)}. \tag{4.7}$$

Thus, the Solow index can also be interpreted as a measure of the change in output divided by a measure of the change in an aggregate input, where the input aggregation rule is the current-period production function in (4.6) and the base-period production function in (4.7).[4]

[4]Solow's assumption of first-degree homogeneity of the production function is not

Following up on a suggestion of Diewert [1992a], Bjurek [1996] exploited the CCD theory of Malmquist indices of aggregate input and output change (based respectively on Malmquist-Shephard input and output distance functions) to develop multiple-output (as well as multiple-input) productivity indices analogous to (4.6) and (4.7).[5] Diewert attributed the ideas behind this index formulation to Moorsteen [1961] and (maybe) Hicks [1961], and the *Hicks-Moorsteen productivity index* assignation has stuck.[6] These indices are exposited in Section 4.4.

The CCD and Bjurek papers are essential to understanding most of today's research on multiple-output productivity, and the remainder of the chapter covers extensions and refinements of the Malmquist and Hicks-Moorsteen indices. The possibility of technological inefficiency—production below the frontier—is implicit in the CCD-Bjurek framework, since the Malmquist-Shephard input and output distance functions serve as measures of inefficiency as well as shifts in the production frontier. Exploiting this dual use of the distance functions, Färe, Grosskopf, Norris, and Zhang [1994] decompose the Malmquist index into two components: technological change (shifts in the production frontier) and efficiency change (movements toward or away from the frontier). This decomposition is discussed in Section 4.5.

Much—perhaps most—of the theoretical research on productivity indices in recent years has been directed at the incorporation of additional decomposition components, like scale effects and input- or output-mix effects; this literature is also reviewed in Section 4.5.

The Malmquist and Hicks-Moorsteen indices, owing to their radial structure, are not applicable to the measurement of productivity in the full space of inputs and outputs.[7] Section 4.6 describes two non-radial indices: the hyperbolic index and the (Luenberger [1992]) directional-distance index.

Dual productivity indices (employing cost and revenue functions) and aggregation of productivity indices across economic units are discussed in

---

needed for any of the foregoing concepts to be well defined, but without it (4.6) and (4.7) fail to satisfy the proportionality property described below in Subsection 4.2.5. (The Hicks-Moorsteen generalization of the Solow index does satisfy this property, but the Malmquist index does not unless the technology satisfies constant returns to scale.)

[5] And (4.2) as well.

[6] Though Balk [2015] proposes renaming it the "Moorsteen-Bjurek index".

[7] Nor are they appropriate for productivity measurement in input or output space in the presence of "bads" like pollution; see Chapter 8 of this volume.

Sections 4.7 and 4.8, respectively.

In the concluding comments in Section 4.9, I take liberties—for the most part resisted in earlier sections—to offer some evaluative comments, particularly with respect to the comparison of the Malmquist and Hicks-Moorsteen approaches.

4.2 Preliminaries: Technological Change and Productivity Indexes

4.2.1 Technologies

Although I frequently refer to the unit of analysis as a "firm," the theory is applicable to any type of economic unit, even an aggregate (national) economy. The firm produces a vector of outputs $y \in \mathbf{R}^m_+$ using a vector of inputs $x \in \mathbf{R}^n_+$. To avoid the nuisance of dealing with null vectors, some of our functions map from production space with the origins of input and output space expunged[8]: $\dot{\mathbf{R}}^{n+m}_+ = \mathbf{R}^n_+ \backslash \{0^{[n]}\} \times \mathbf{R}^m_+ \backslash \{0^{[m]}\} =: \dot{\mathbf{R}}^n_+ \times \dot{\mathbf{R}}^m_+$.

The firm's production is constrained by a (known) technology—the set of technologically feasible production vectors,

$$T \;=\; \Big\{\langle x, y\rangle \in \mathbf{R}^{n+m}_+ \mid x \text{ can produce } y\Big\}.$$

Given the technology T, the output-possibility set for input vector x is[9]

$$P(x, T) = \Big\{y \in \mathbf{R}^m_+ \mid \langle x, y\rangle \in T\Big\}$$

and the input-requirement set for output vector y is

$$L(y, T) = \Big\{x \in \mathbf{R}^n_+ \mid \langle x, y\rangle \in T\Big\}.$$

[8]This problem can be dealt with in other ways (see footnote 12 in Russell [1998]), but given the unimportance of the shutdown condition, I think the simple approach of purging the null vector from the domain is best.

[9]It is not standard in the literature to make explicit the dependence of the input-requirement and production-possibility sets on the specification of the technology, but it is convenient to do so when we contemplate productivity changes in the face of changes in the technology over time or differences of the technology among production units.

Clearly, $\langle x, y\rangle \in T \iff x \in L(y,T) \iff y \in P(x,T)$.

We restrict the set of technologies, denoted $\mathcal{T}$, to those that are closed and satisfy free input disposability ($L(y,T) + \mathbf{R}^n_+ = L(y,T)$ for all $y \in \mathbf{R}^m_+$) and free output disposability ($P(x,T) = (P(x,T) - \mathbf{R}^m_+) \bigcap \mathbf{R}^m_+$ for all $x \in \mathbf{R}^n_+$). As T is closed for all $T \in \mathcal{T}$, so are the slices, $L(y,T)$ and $P(x,T)$, for all $\langle x, y\rangle \in \mathbf{R}^{n+m}_+$.

The isoquant for output $y \in \dot{\mathbf{R}}^m_+$ is given by

$$I(y,T) = \{x \in L(y,T) \mid \lambda x \notin L(y) \ \forall\ \lambda < 1\},$$

and the production possibility surface for input $x \in \mathbf{R}^n_+$ is

$$\Gamma(x,T) = \{y \in P(x,T) \mid \lambda y \notin P(x) \ \forall\ \lambda > 1\}.$$

4.2.2 Distance Functions

An essential building block of multiple-output (and, of course, multiple-input) productivity analysis is the "distance function," independently introduced into the economics literature,[10] in different contexts and with different objectives, by Debreu [1951], Malmquist [1953], and Shephard [1953]. Malmquist's context is the closest to the thrust of this chapter (even though his objective was the construction of consumer cost-of-living indices rather than productivity indices).

The input distance function, $D_I : \mathbf{N} \times \mathcal{T} \to \mathbf{R}_{++}$, maps from a subset of production space, $\mathbf{N} = \left\{\langle x, y\rangle \in \dot{\mathbf{R}}^{n+m}_+ \mid L(y,T) \neq \emptyset\right\}$, and the set of allowable technologies into the positive real line and is defined by

$$D_I(x,y,T) = \max\left\{\lambda > 0 \mid x/\lambda \in L(y,T)\right\}.$$

Thus, the input distance function is defined as the maximal (proportional) radial contraction (or minimal radial expansion) of a given input vector consistent with technological feasibility of production of a given output vector.[11]

[10] Mathematicians refer to it as the "gauge function"; "distance function" is a distinctive (and different) concept in real analysis.

[11] The input distance function in the Solow set-up is given by $D^S_I(x,y) = \max\left\{\lambda > 0 \mid y \leq f(x/\lambda)\right\}$. If f is homogenous of degree 1, $D^S_I(x,y) = f(x)/y$.

Thus, the input distance function can be characterized as a notion of "radial distance" of an input-quantity vector from the frontier of the technology (the isoquant in input space)—a characterization that, as we shall see, is evocative in the construction of productivity indices.

With our (parsimonious) restrictions on the technology (closedness and free input disposability), D_I is well-defined, homogeneous of degree one and non-decreasing in x, and non-increasing in y for all $\langle x, y, T\rangle \in \mathbf{N} \times \mathcal{T}$. Moreover, $x \in L(y,T) \iff D_I(y,x,T) \geq 1$, and $x \in I(y,T) \iff D_I(x,y,T) = 1$, so that, for any $y \in \dot{\mathbf{R}}^m_+$, $L(y,T)$ is recovered from D_I by

$$L(y,T) = \{x \in \mathbf{R}^n_+ \mid D_I(x,y,T) \geq 1\}$$

and $I(y,T)$ is recovered from D_I by

$$I(y,T) = \{x \in \mathbf{R}^n_+ \mid D_I(x,y,T) = 1\}.$$

Thus, the input distance function also serves as a functional representation of the technology.

The output distance function, $D_O : \mathbf{N} \times \mathcal{T} \to \mathbf{R}_+$, is similarly defined by

$$D_O(x,y,T) = \min\left\{\lambda > 0 \mid y/\lambda \in P(x,T)\right\};$$

i.e., as the minimal (proportional) radial contraction (or maximal radial expansion) of a given output vector consistent with technological feasibility of production for a given input vector. Thus, the output distance function is the "radial distance" of an output-quantity vector from the frontier of the technology (the production possibility surface in output space).[12]

With our restrictions on the technology (closedness, free output disposability, and boundedness of $P(x,T)$), D_O is well-defined, homogeneous of degree one and non-decreasing in y, and non-increasing in x for all $\langle x, y\rangle \in \dot{\mathbf{R}}^{n+m}_+$. Moreover, $y \in P(x,T) \iff D_O(x,y,T) \leq 1$, so that, for any $x \in \dot{\mathbf{R}}^n_+$, $P(x,T)$ is recovered by

$$P(x,T) = \left\{y \in \mathbf{R}^m_+ \mid D_O(x,y,T) \leq 1\right\}$$

and $\Gamma(x)$ is recovered by

$$\Gamma(x,T) = \left\{y \in \mathbf{R}^m_+ \mid D_O(x,y,T) = 1\right\},$$

[12]The output distance function in the Solow set-up is given by $D^S_O(x,y,T) = \min\{\lambda > 0 \mid y/\lambda \leq f(x)\} = y/f(x)$.

indicating that the output distance function is a functional representation of the technology.

Finally, it is easy to see that each of these distance functions is independent of units of measurement.[13]

4.2.3 Technological Efficiency

A firm is input inefficient at time t if $D_I(x^t, y^t, T^t) > 1$ and output inefficient if $D_O(x^t, y^t, T^t) < 1$. Moreover, with restriction of the distance-function domains to feasible production vectors $\langle x^t, y^t \rangle \in T^t$, $1/D_I(x^t, y^t, T^t) =: E_I(x^t, y^t, T^t)$ and $D_O(x^t, y^t, T^t) =: E_O(x^t, y^t, T^t)$ are, respectively, *(Debreu-Farrell) input and output efficiency indices* (Debreu [1951] and Farrell [1957]). Each measures the radial distance of the quantity vector from the frontier and, for technologically feasible production vectors, maps into the $(0, 1]$ interval.

4.2.4 Technological Change

The essence of technological comparisons across periods (or across production units) is comparison of production possibilities, as reflected by production possibility sets or input requirement sets, under counterfactual assumptions about input availability or output requirement.

The production possibility set at the *current-period* input vector using the *base-period* technology is $P(x^c, T^b)$, and the production possibility set at the base-period input vector using the current-period technology is $P(x^b, T^c)$. The corresponding production possibility surfaces are $\Gamma(x^c, T^b)$ and $\Gamma(x^b, T^c)$.

If

$$P(x^c, T^b) \subset P^c(x^c, T^c) \quad \wedge \quad \Gamma(x^c, T^b) \cap \Gamma(x^b, T^c) = \emptyset, \tag{4.8}$$

we say that, measuring qualitatively in output space and normalizing on the current-period input vector, the technology of the economic unit unambiguously (globally) improved between b and c. This normalization is arbitrary, and we could just as well normalize on the base-period input vector, in which

[13]For proofs, discussions, and illustrations of the properties of input and output distance functions, see Färe and Primont [1995] and Russell [1998].

case the technology of the economic unit unambiguously improved if

$$P(x^b, T^b) \subset P(x^b, T^c) \quad \wedge \quad \Gamma(x^c, T^b) \cap \Gamma(x^b, T^c) = \emptyset. \tag{4.9}$$

(These characterizations of unambiguous technological improvement can similarly be expressed in terms of (shrinking) input requirement sets at stipulated output vectors.)

These criteria for global technological change are overly strong: a more reasonable requirement is to normalize on the output quantity vector as well—*i.e*, require only that the production possibility set expand in a neighborhood of the output quantity vector, y^c in (4.8) or y^b in (4.9). This is the approach adopted by CCD in their pathbreaking analysis of productivity measurement, entailing quantitative as well as qualitative measurement using the output distance function as a measure of the radial distance of an output vector in one period from the production possibility surface of the technology in the other period.

The qualitative characterization of technological change in input space is analogous. The input requirement set at the *current-period* output vector using the *base-period* technology is $L(y^c, T^b)$, and the input requirement set at the base-period output vector using the current-period technology is $L(y^b, T^c)$. The corresponding isoquants are $I(y^c, T^b)$ and $I(y^b, T^c)$.

Under alternative normalizations, the technology of the economic unit unambiguously (globally) improves between b and c if $L(y^c, T^c) \subset L(y^c, T^b)$ or $L(y^b, T^c) \subset L(y^b, T^b)$. More reasonably, we can say that there has been technological progress (in the appropriate normalization) if the input requirement set expands (toward the origin) in a neighborhood of x^c or, alternatively, x^b.

4.2.5 Productivity Indices

At the most abstract, generic (even austere) level, the definition of a multiple-output productivity index is a mapping, $\Pi : \mathbf{R}_+^{2(n+m)} \to \mathbf{R}_{++}$, with image $\Pi(x^b, x^c, y^b, y^c)$. Without some imposed structure or required set of properties, however, this concept is close to vacuous. The two main approaches to adding structure are the axiomatic, or test, approach (requiring that the index satisfy certain properties, like monotonicity) and the economic approach (tying the index to economic or technological constructs). (These approaches are complementary.)

The desirable properties of productivity indices include the following:[14]

(I) *Identity*: $\langle x^b, y^b\rangle = \langle x^c, y^c\rangle \Longrightarrow \Pi(x^b, x^c, y^b, y^c) = 1$.

(M) *Monotonicity*: Productivity is nondecreasing in current-period output and base-period inputs and nonincreasing in current-period inputs and base-period outputs:

$$\langle x^b, -x^c, -y^b, y^c\rangle \geq \langle \hat{x}^b, -\hat{x}^c, -\hat{y}^b, \hat{y}^c\rangle \\ \Longrightarrow \Pi(x^b, x^c, y^b, y^c) \geq \Pi(\hat{x}^b, \hat{x}^c, \hat{y}^b, \hat{y}^c).$$

(UI) *Invariance with respect to units of measurement (commensurability)*:

$$\Pi(x^b K_x, x^c K_x, y^b K_y, y^c K_y) = \Pi(x^b, x^c, y^b, y^c),$$

for arbitrary $n \times n$ and $m \times m$ positive diagonal (unit transformation) matrices, K_x and K_y.

(P) *Proportionality*: Π is homogeneous of degree 1 in y^c and homogeneous of degree -1 in x^c (hence homogeneous of degree zero in $\langle x^c, y^c\rangle$).

(T) *Transitivity*: For any three periods, b, c, and d, the product of the measured productivity changes from period b to period c and from period c to period d is equal to the productivity change from period b to period d:

$$\Pi(x^b, x^c, y^b, y^c) \cdot \Pi(x^c, x^d, y^c, y^d) = \Pi(x^b, x^d, y^b, y^d).$$[15]

Linkage to technological or economic concepts requires a more constructive approach. To this end, we expand the domain to encompass the space of technologies $\mathcal{T}$.[16] The extended productivity index (in a slight abuse of notation), $\Pi : \mathbf{R}_+^{2(n+m)} \times \mathcal{T}^2 \to \mathbf{R}_{++}$, now has the image $\Pi(x^b, x^c, y^b, y^c, T^b, T^c)$.

[14]For thorough examinations of properties of productivity indices, see Diewert [1992b] and Balk [1998].

[15]This condition is equivalent to *circularity*, $\Pi(x^b, x^c, y^b, y^c) \cdot \Pi(x^c, x^d, y^c, y^d) \cdot \Pi(x^d, x^b, y^d, y^b) = 1$, if (and only if) the identity condition holds, but the two conditions are typically used interchangeably under the implicit assumption that the latter condition holds.

[16]Dependence on economic variables, like prices, is briefly discussed in Subsection 4.7.

In many specific cases the index is invariant with respect to changes in some quantity vectors or one of the technologies.

Application of the above axioms to indices with the enhanced domain is straightforward. (One could add axioms related to the technologies, but this approach, to my knowledge, has not been explored.)

4.3 Malmquist Productivity Indices

Caves, Christensen, and Diewert [1982] (CCD) imposed structure by a natural linkage to the technology using the distance function as a representation of the technology and as a notion of distance from a quantity vector in one period to a technological frontier in another. They proposed four basic productivity indices predicated on alternative normalizations with respect to the choice of the base period or the current period for the reference technology and for the production vector. These indices are defined as follows (where the relation $\overset{\text{eff}}{=}$ holds only under the assumption of technological efficiency):[17]

Output-based, technology b-based Malmquist productivity index:

$$\begin{aligned}\Pi_O^M(x^b, x^c, y^b, y^c, T^b) &= \frac{D_O(x^c, y^c, T^b)}{D_O(x^b, y^b, T^b)} \overset{\text{eff}}{=} D_O(x^c, y^c, T^b) \\ &= \min\left\{\lambda > 0 \mid y^c/\lambda \in P(x^c, T^b)\right\} =: \bar{\Pi}_O^M(x^c, y^c, T^b). \qquad (4.10)\end{aligned}$$

Output-based, technology c-based Malmquist productivity index:

$$\begin{aligned}\Pi_O^M(x^b, x^c, y^b, y^c, T^c) &= \frac{D_O(x^c, y^c, T^c)}{D_O(x^b, y^b, T^c)} \overset{\text{eff}}{=} \frac{1}{D_O(x^b, y^b, T^c)} \\ &= \left(\min\left\{\lambda > 0 \mid y^b/\lambda \in P(x^b, T^c)\right\}\right)^{-1} =: \bar{\Pi}_O^M(x^b, y^b, T^c). \quad (4.11)\end{aligned}$$

Input-based, technology b-based Malmquist productivity index:

$$\Pi_I^M(x^b, x^c, y^b, y^c, T^b) = \frac{D_I(x^b, y^b, T^b)}{D_I(x^c, y^c, T^b)} \overset{\text{eff}}{=} \frac{1}{D_I(x^c, y^c, T^b)}$$

[17] Comprehension of the concepts discussed throughout this chapter would be enhanced by drawing the diagrams that take up too much space to include in this survey. For diagrammatic expositions of many of these concepts, see Russell [1998].

$$= \left(\max \left\{ \lambda > 0 \mid x^c/\lambda \in L(y^c, T^b) \right\} \right)^{-1} =: \bar{\Pi}_I^M(x^c, y^c, T^b). \quad (4.12)$$

Input-based, technology c-based Malmquist productivity index:

$$\begin{aligned} \Pi_I^M(x^b, x^c, y^b, y^c, T^c) &= \frac{D_I(x^b, y^b, T^c)}{D_I(x^c, y^c, T^c)} \overset{\text{eff}}{=} D_I(x^b, y^b, T^c) \\ &= \max \left\{ \lambda > 0 \mid x^b/\lambda \in L(y^b, T^c) \right\} =: \bar{\Pi}_I^M(x^b, y^b, T^c). \end{aligned} \quad (4.13)$$

Thus, each of these productivity indices is defined, in the first instance, as a ratio of radial distances to an (input or output) frontier for base-period and current-period quantities and for a common technology.[18] Under the assumption that production units operate (efficiently) on the production frontier, $D_O(x^b, y^b, T^b) = D_I(x^b, y^b, T^b) = 1$ for $t = b, c$, so that the indices simplify to single distance functions evaluated at a mixture of a production vector for one period and a technology for the other period.

The (output-based) productivity index normalized on the base-period technology, defined in (4.10), measures the maximal radial contraction of the current-period output vector needed to place the contracted vector in the production possible set for current-period input, using the base-period technology. Of course, if $y^c \in P(x^c, T^b)$, it must be expanded radially to the frontier. Clearly, $\bar{\Pi}_O^M(x^c, y^c, T^b) > 1$ if and only if there is technological progress—expansion of the frontier—in the neighborhood of the current-period production vector; conversely, $\bar{\Pi}_O^M(x^c, y^c, T^b) < 1$ if and only if the frontier of the production possibility set has receded in this neighborhood, evincing technological retardation. Note that this index, under technological efficiency, is precisely a generalization to multiple outputs (and to non-homothetic technologies) of Solow's measure of technical progress in the scalar-output case (equation (4.5) above).[19]

The calculation of the (output-based) productivity index normalized on the current-period technology in (4.11) generates the inverse of the minimal expansion of the base-period output vector required to reach the frontier of the current-period production possibility set for base-period input vector.

[18] I follow convention in normalizing on either the base-period or current-period technology, but the Malmquist indices could in principle be defined with respect to any technology in $\mathcal{T}$. *Cf.* footnote 22

[19] Use footnote 12 to obtain $D_O(x^c, y^c, T^b) = y^c / f^b(x^c)$.

$\bar{\Pi}_O^M(x^b, y^b, T^c) > 1$ if and only if there has been expansion of the frontier of production possibility set in the neighborhood of the base-period production vector and $\bar{\Pi}_O^M(x^b, y^b, T^b) < 1$ if and only if the frontier of the production possibility set has receded in this neighborhood. This index is a generalization of the scalar-output measure of technological change in (4.4) above.[20]

The (input-based) productivity index normalized on the base-period technology, defined in (4.12), measures the inverse of the maximal radial expansion of the current-period input vector needed to place the expanded vector in the input-requirement set for current-period output, using the base-period technology. Clearly, $\bar{\Pi}_I^M(x^c, y^c, T^b) > 1$ if and only if there is technological progress—expansion of the isoquant toward the origin—in the neighborhood of the current-period production vector. This index is an alternative generalization of the scalar-output measure of technological change in (4.5) above.[21]

Interpretation of the (input-based) productivity index normalized on the current-period technology, defined in (4.13) is similar, and again $\bar{\Pi}_I^M(x^b, y^b, T^c) > 1$ if and only if technological progress has lowered the isoquant in the neighborhood of base-period quantities.

These four indices, in general, provide different implications about productivity change, even possibly about the direction of change. The delineations among them is arbitrary enough to suggest that there is unlikely to be a compelling criterion for choosing any one. One suggestion, inspired by Irving Fisher's recommendation for dealing with arbitrary normalizations in the construction of index numbers is to take geometric averages over current-period and base-period constructions. He referred to these confluences as "ideal" indices. In this spirit, Diewert [1992a] suggested "Fisher-ideal" indices for input-oriented and for output-oriented productivity indices; they are defined as follows:

Malmquist "ideal" output-based productivity index:

$$\begin{aligned} \Pi_O^{MI}(x^b, x^c, y^b, y^c, T^b, T^c) &= \left(\Pi_O^M(x^b, x^c, y^b, y^c, T^b) \cdot \Pi_O^M(x^b, x^c, y^b, y^c, T^c)\right)^{1/2} \\ &\overset{\text{eff}}{=} \left(\frac{D_O(x^c, y^c, T^b)}{D_O(x^b, y^b, T^c)}\right)^{1/2}. \end{aligned} \tag{4.14}$$

[20] Use footnote 12 to obtain $1/D_O(x^b, y^b, T^c) = f^c(x^b)/y^b$.
[21] Use footnote 11 to obtain $1/D_I(x^c, y^c, T^b) = y^c/f^b(x^c)$.

Malmquist "ideal" input-based productivity index:

$$\begin{aligned}\Pi_I^{MI}(x^b, x^c, y^b, y^c, T^b, T^c) &= \left(\Pi_I^M(x^b, x^c, y^b, y^c, T^b) \cdot \Pi_I^M(x^b, x^c, y^b, y^c, T^c)\right)^{1/2} \\ &\underset{=}{\text{eff}} \left(\frac{D_I(x^b, y^b, T^c)}{D_I(x^c, y^c, T^b)}\right)^{1/2}. \end{aligned} \tag{4.15}$$

The properties of the input and output distance functions, described above, endow each of these indices with the identity (I), monotonicity (M), and unit-invariance (UI) properties. On the other hand, these indices satisfy neither proportionality (P) nor transitivity (T) on their domains.[22] In the next section I describe an index that does satisfy proportionality (though not transitivity).

4.4 Hicks-Moorsteen Indices

Recall from equation (4.2) that Solow technical change for a single output can be interpreted as a ratio of total factor productivity in two periods. The various Malmquist indices are indisputably measures of technological change—that is, shifts in the frontier of the technology—but in general they do not have the interpretation as ratios of aggregate productivity in the two periods. Nor can they be interpreted in general as ratios of (aggregate) output change to (aggregate) input change, unlike the Solow single-output index in the form given by identity (4.6) or (4.7).

Building on the construction in CCD (and following a suggestion of Diewert [1992a]), Bjurek [1996] developed multiple-output indices analogous to (4.6) and (4.7). The key building blocks for these indices are essentially Malmquist distance functions evaluated at intertemporal mixtures of production vectors (base-period input vector and current period output vector, or *vice versa*). Define the *Malmquist quantity indices* as follows (where,

[22] See Färe and Grosskopf [1996]. Input-oriented (output-oriented) Malmquist indices satisfy (P) if the distance function is homogeneous of degree -1 in output quantities (homogeneous of degree -1 in input quantities). To get around the non-transitivity problem, Berg, Førsund, and Jansen [1992] and Pastor and Lovell [2005], following a suggestion of Diewert [1987], propose indices defined on a particular technology (the first-period technology or the union of all technologies). This approach is critiqued by Balk and Althin [1996], who then propose a more elaborate construction to impose transitivity.

again, the last equation in each case holds only under the assumption that firms operate on the technological frontier):

Technology b-based Malmquist output index:

$$Q_O^M(x^b, y^b, x^c, y^c, T^b) = \frac{D_O(x^b, y^c, T^b)}{D_O(x^b, y^b, T^b)} \overset{\text{eff}}{=} D_O(x^b, y^c, T^b). \tag{4.16}$$

Technology c-based Malmquist output index:

$$Q_O^M(x^b, y^b, x^c, y^c, T^c) = \frac{D_O(x^c, y^c, T^c)}{D_O(x^c, y^b, T^c)} \overset{\text{eff}}{=} \frac{1}{D_O(x^c, y^b, T^c)}. \tag{4.17}$$

Technology b-based Malmquist input index:

$$Q_I^M(x^b, y^b, x^c, y^c, T^b) = \frac{D_I(x^c, y^b, T^b)}{D_I(x^b, y^b, T^b)} \overset{\text{eff}}{=} D_I(x^c, y^b, T^b). \tag{4.18}$$

Technology c-based Malmquist input index:

$$Q_I^M(x^b, y^b, x^c, y^c, T^c) = \frac{D_I(x^c, y^c, T^c)}{D_I(x^b, y^c, T^c)} \overset{\text{eff}}{=} \frac{1}{D_I(x^b, y^c, T^c)}. \tag{4.19}$$

Identity (4.16) yields a measure of the (aggregate) output change between the base period b and the current period c. Specifically, it is the radial reduction of current-period output required to place the contracted output vector in the production possibility set for base-period input vector using the base-period technology. As such, it can be interpreted as the radial distance of the current-period output vector from the base-period production possibility set. Obviously, $Q_O(x^b, y^b, x^c, y^c, T^b) > 1$ if (and only if) $y^c \notin P(x^b, T^b)$ (*i.e.*, the current-period output vector lies above the production possibility surface for base-period technology and input vector), in which case we say that output increased.

Identity (4.17), on the other hand, yields the maximal expansion of the base-period output vector consistent with the expanded output vector remaining in the production possibility set for current-period input vector using the current-period technology. As such, it is a measure of the radial distance of the base-period output vector from the current period technological frontier. Clearly, $Q_O(x^b, y^b, x^c, y^c, T^c) > 1$ if and only if y^b is below the current-period output frontier, in which case we say that output increased.

The input indices in (4.18) and (4.19) have analogous interpretations, in terms of radial distances of input vectors in one period to isoquants of the alternative period.

The *Hicks-Moorsteen productivity indices* can now be defined as the ratios of output changes to input changes, normalized alternatively on base-period and current-period technologies:

$$\Pi^{HM}(x^b, x^c, y^b, y^c, T^b) = \frac{Q_O^M(x^b, x^c, y^b, y^c, T^b)}{Q_I^M(x^b, x^c, y^b, y^c, T^b)} \underset{=}{\text{eff}} \frac{D_O(x^b, y^c, T^b)}{D_I(x^c, y^b, T^b)} \tag{4.20}$$

and

$$\Pi^{HM}(x^b, x^c, y^b, y^c, T^c) = \frac{Q_O^M(x^b, x^c, y^b, y^c, T^c)}{Q_I^M(x^b, x^c, y^b, y^c, T^c)} \underset{=}{\text{eff}} \frac{D_I(x^b, y^c, T^c)}{D_O(x^c, y^b, T^c)}. \tag{4.21}$$

The Hicks-Moorsteen index can also be interpreted as a multiple-output generalization of the total-factor-productivity version of the Solow index, equation (4.2). Re-write, *e.g.*, the first equality in (4.20) as

$$\Pi^{HM}(x^b, x^c, y^b, y^c, T^b) = \frac{D_O(x^b, y^c, T^b)/D_I(x^c, y^b, T^b)}{D_O(x^b, y^b, T^b)/D_I(x^b, y^b, T^b)}. \tag{4.22}$$

The numerator is interpreted as the period-c ratio of a (static) output quantity index normalized on base-period input quantity and technology to an input quantity index normalized on base-period input quantity vector and technology. The denominator is similarly interpreted as an aggregate output/input ratio in the base period b.

Finally, in the absence of a compelling criterion for choosing between the base-period and current-period normalizations, Bjurek suggests the "Fisher ideal" amalgamation to obtain the *Hicks-Moorsteen "ideal" productivity index*:

$$\Pi^{HMI}(x^b, x^c, y^b, y^c, T^b, T^c) = \frac{Q_O^I(x^b, x^c, y^b, y^c, T^b, T^c)}{Q_I^I(x^b, x^c, y^b, y^c, T^b, T^c)}, \tag{4.23}$$

where

$$Q_O^I(x^b, x^c, y^b, y^c, T^b, T^c) = \left(Q_O^M(x^b, x^c, y^b, y^c, T^b) \ \cdot \ Q_O^M(x^b, x^c, y^b, y^c, T^c) \right)^{1/2}$$

and

$$Q_I^I(x^b, x^c, y^b, y^c, T^b, T^c) = \Big(Q_I^M(x^b, x^c, y^b, y^c, T^b) \ \cdot \ Q_I^M(x^b, x^c, y^b, y^c, T^c)\Big)^{1/2}$$

are *Fisher "ideal" quantity indices*.

A comparison of the Malmquist and Hicks-Moorsteen indices is instructive. In a nutshell, as first emphasized by Grifell-Tatjé and Lovell [1995], the former is a measure of *technological change* (shift in the production frontier) while the latter is a (broader) measure of the change in *total-factor productivity* (incorporating the effects of movement along the frontier as well as the shift of the frontier). Maintaining technological efficiency, each of the four Malmquist indices (4.10)–(4.13) measures productivity as the radial shift of the frontier (in either input or output space) in the neighborhood of a *given production vector*. The Hicks-Moorsteen indices, (4.20) and (4.21), on the other hand, incorporate the effects of changes in the production vector—most notably, scale effects[23]—as well as shifts in the technological frontier.

This comparison suggests that the difference between the two types of indices vanishes if productivity is invariant with respect to shifts of the production vector along the surface of the technology. Indeed, this fact has been proved by Färe, Grosskopf, and Roos [1996]. A straightforward generalization of their finding is as follows: if and only if, for all t, the technology satisfies (a) input and output homotheticity (isoquants and production possibility surfaces are radial transformations of one another)[24] and (b) constant returns to scale ($D_O(\lambda x^b, y^b, T^b) = (1/\lambda) \cdot D_O(x^b, y^b, T^b)$ for all $\langle x^b, y^b \rangle \in T^b$), the four Malmquist indices (4.10)–(4.13) and the two Hicks-Moorsteen indices (4.20) and (4.21) are identical (cf. Balk [1998, 112–114]).

[23] But also output mix effects; see Balk [2001] and the discussion of the decomposition of a productivity change into its separate components in Section 4.5 below.

[24] Färe and Primont [1995, 69–72] showed that this property, which they called *inverse homotheticity*, is equivalent to $D_O(x, y, T) = D_O(\bar{x}, y, T)/\Psi\big(D_I(x, \bar{y}, T)$ for an arbitrary $\langle \bar{x}, \bar{y} \rangle \in \dot{\mathbf{R}}_+^{n+m}$ and for some strictly monotonic function Ψ.

4.5 Decompositions of Productivity Growth

4.5.1 Inefficiency and Technological Change

Recall that CCD initially defined the four (output- and input-based, technology-based) Malmquist indices as ratios of distance functions. But under the assumption that production units operate (efficiently) on the production frontier, the indices simplify to the expressions to the right of the relation $\overset{\text{eff}}{=}$ in (4.10)–(4.13). Färe, Grosskopf, Norris and Zhang [1994] (FGNZ) elaborated on the CCD framework by allowing firms to operate inefficiently, below the production possibility surface for the extant input vector or above the isoquant for the extant output vector.

The acknowledgement that economic units might operate less than fully efficiently opens up the possibility of decomposing the Malmquist index into two components: changes in the technology (shifts of the frontier) and changes in efficiency (radial distance from the frontier). FGNZ executed this idea by rewriting the initial identity in (4.10)–(4.13) in terms of efficiency indices (see Subsection 4.2.3) as follows:

$$\Pi_O^M(x^b, x^c, y^b, y^c, T^b) = \frac{D_O(x^c, y^c, T^b)}{D_O(x^c, y^c, T^c)} \frac{E_O(x^c, y^c, T^c)}{E_O(x^b, y^b, T^b)}, \tag{4.24}$$

$$\Pi_O^M(x^b, x^c, y^b, y^c, T^c) = \frac{D_O(x^b, y^b, T^b)}{D_O(x^b, y^b, T^c)} \frac{E_O(x^c, y^c, T^c)}{E_O(x^b, y^b, T^b)}, \tag{4.25}$$

$$\Pi_I^M(x^b, x^c, y^b, y^c, T^b) = \frac{D_I(x^c, y^c, T^c)}{D_I(x^c, y^c, T^b)} \frac{E_I(x^c, y^c, T^c)}{E_I(x^b, y^b, T^b)}, \tag{4.26}$$

and

$$\Pi_I^M(x^b, x^c, y^b, y^c, T^c) = \frac{D_I(x^b, y^b, T^c)}{D_I(x^b, y^b, T^b)} \frac{E_I(x^c, y^c, T^c)}{E_I(x^b, y^b, T^b)}. \tag{4.27}$$

In each case, the Malmquist index allowing for the possibility of inefficient production is the multiple of two indices: the two ratios on the right-hand sides. The first ratio is an index of technical change: the radial shift, in either input space or output space, of the production frontier in the neighborhood of either the base-period or the current-period production vector. The second is an index of the change in efficiency, oriented alternatively to output space or input space.

As in the case of efficient production, the output-oriented and input-oriented Malmquist "ideal" indices are defined as the geometric means taken over the indices normalized on the base period and the current period:

$$\begin{aligned}
&\Pi_O^{MI}(x^b, x^c, y^b, y^c, T^b, T^c) \\
&\quad = \left(\Pi_O^M(x^b, x^c, y^b, y^c, T^b) \cdot \Pi_O^M(x^b, x^c, y^b, y^c, T^c)\right)^{1/2} \\
&\quad = \left(\frac{D_O(x^c, y^c, T^b)}{D_O(x^c, y^c, T^c)} \cdot \frac{D_O(x^b, y^b, T^b)}{D_O(x^b, y^b, T^c)}\right)^{1/2} \frac{E_O(x^c, y^c, T^c)}{E_O(x^b, y^b, T^b)}
\end{aligned}$$

and

$$\begin{aligned}
&\Pi_I^{MI}(x^b, x^c, y^b, y^c, T^b, T^c) \\
&\quad = \left(\Pi_I^M(x^b, x^c, y^b, y^c, T^b) \cdot \Pi_I^M(x^b, x^c, y^b, y^c, T^c)\right)^{1/2} \\
&\quad = \left(\frac{D_I(x^c, y^c, T^c)}{D_I(x^c, y^c, T^b)} \cdot \frac{D_I(x^b, y^b, T^c)}{D_I(x^b, y^b, T^b)}\right)^{1/2} \frac{E_I(x^c, y^c, T^c)}{E_I(x^b, y^b, T^b)}.
\end{aligned}$$

Thus, as originally formulated, the CCD indices capture efficiency change as well as technological change, and FGNZ capitalized on this formulation to implement a binary decomposition of productivity change into these two components. The success of this decomposition, particularly in empirical studies, provided an impetus to refine the decomposition into additional contributing factors. These efforts are summarized in the next subsection.

4.5.2 Returns to Scale and Output Mix

As noted above, the Malmquist productivity index does not encompass the effects of returns to scale (RTS) as quantities change. Thus, it is not surprising that efforts to incorporate RTS into the decomposition of the Malmquist index[25] led instead to revisions of the index itself

The first of these revisions is a "Malmquist index" defined, not on the actual technology, but instead on a virtual constant-returns-to-scale (CRS) technology, a conical envelopment of the actual technology. Among the many

[25]See Färe, Grosskopf, Norris, and Zhang [1994], Ray and Desli [1997], Grifell-Tatjé and Lovell [1999], and Wheelock and Wilson [1999]. These contributions are reviewed by Balk [2001], Grosskopf [2003], and Zofio [2007].

papers directed at this issue, perhaps the "constructive" approach—starting with individual components of a productivity index and then aggregating over these components to construct a consistent overall productivity index—best illustrates this progression. A good example is the contribution of Balk [2001], who referred to it as the "bottom up" approach (as opposed to the "top down" approach begun by FGNZ).

Define the conical envelopment of the technology T by

$$\mathcal{C}(T) = \{\langle \lambda x, \lambda y\rangle \in \dot{\mathbf{R}}_+^{n+m} \mid \langle x, y\rangle \in T \ \wedge \ \lambda > 0\}$$

and restrict the set of allowable technologies to those for which $\mathcal{C}(T)$ is a proper subset of $\mathbf{R}_+^{n+m}$ (thus excluding, *e.g.*, technologies with globally increasing returns to scale and Cobb-Douglas-like technologies where some marginal product goes to infinity as the quantity goes to zero). The cone $\mathcal{C}(T)$ is a virtual technology—informally, the "smallest" CRS technology containing T.

Roughly speaking, scale efficiency at a point in production space is measured as the "proportional distance" from the actual technological frontier to the conical (envelopment) technological frontier in either input or output space. I stick here to the output notion; the alternative approach in input space can be found in Balk [2001]. Output scale efficiency is thus defined as the ratio of output efficiency defined on the virtual technology divided by output efficiency defined on the true technology:

$$S_O(x, y, T) = \frac{E_O(x, y, \mathcal{C}(T))}{E_O(x, y, T)} = \frac{D_O(x, y, \mathcal{C}(T))}{D_O(x, y, T)}.$$

Note that $S_O(x, y, T) \leq 1$ and, if the technology satisfies constant returns to scale, $\mathcal{C}(T) = T$ and $S(x, y, T) = 1$.

As in the case of the FGNZ decomposition of Malmquist productivity change into technical change and technological efficiency components, there exist more than one path of quantity and technological change along which to measure these components. Again in the face of space constraints, I consider only the path of, first, quantity changes along the base-period technological frontier and, then, shift of the frontier at current-level quantities. (Technological efficiency change is path independent.)

The returns-to-scale index for a change in the input quantity vector from x^b to x^c is the resultant change in the gap between the actual and conical

production frontiers:

$$SI_O(x^b, x^c, y^b, T^b) = \frac{S_O(x^c, y^b, T^b)}{S_O(x^b, y^b, T^b)} = \frac{D_O(x^c, y^b, \mathcal{C}(T^b))/D_O(x^c, y^b, T^b)}{D_O(x^b, y^b, \mathcal{C}(T^b))/D_O(x^b, y^b, T^b)}. \tag{4.28}$$

Interchanging the SW and NE part of the last term yields the interpretation of the index as a measure of the expansion (or contraction) of the *virtual* (conical) production possibility frontier along a ray through y^b resulting from the change in the input vector divided by a measure of the expansion (or contraction) of the *actual* production possibility frontier along a ray through y^b for the same change in the input vector.

As Balk [2001] perspicaciously pointed out, (4.28) does not tell the whole story: as the input vector changes from x^b to x^c, the ray through the output vector also shifts to pass through y^c (unless the technology satisfies output homotheticity). This shift also changes the level of scale efficiency, $S_O(x, y, T)$.[26] Balk refers to this change as the *output-mix effect* and measures it as follows:

$$M_O(x^c, y^c, y^b, T^b) = \frac{D_O(x^c, y^c, \mathcal{C}(T^b)) \ / \ D_O(x^c, y^c, T^b)}{D_O(x^c, y^b, \mathcal{C}(T^b)) \ / \ D_O(x^c, y^b, T^b)}.$$

Combining the returns-to-scale and output-mix effects with the FGNZ technological and efficiency change indices in (4.24),

$$TC_O(x^c, y^c, T^b, T^c) := \frac{D_O(x^c, y^c, T^b)}{D_O(x^c, y^c, T^c)}$$

and

$$EC_O(x^b, x^c, y^b, y^c, T^b, T^c) := \frac{D_O(x^c, y^c, T^c)}{D_O(x^b, y^b, T^b)},$$

we obtain, after some cancellation of terms, the overall productivity index,

$$\begin{aligned} &TC_O(x^c, y^c, T^b, T^c) \cdot EC_O(x^b, x^c, y^b, y^c, T^b, T^c) \\ &\quad \cdot SI_O(x^b, x^c, y^b, T^b) \cdot M_O(x^c, y^c, y^b, T^b) = \frac{D_O(x^c, y^c, \mathcal{C}(T^b))}{D_O(x^b, y^b, \mathcal{C}(T^b))}. \end{aligned}$$

Thus, the four components aggregate to a Malmquist index defined on the virtual (conical) technology $\mathcal{C}(T)$ rather than the actual technology T.

[26] Note that, since $S_O(x, y, T)$ is homogeneous of degree zero in y, its value depends on y only through the location of the ray through y.

The conical-envelopment approach to incorporating returns to scale into the Malmquist index has come under some criticism in recent years. Already noted above is the problem of the index not being defined for all technologies in $\mathcal{T}$. In addition, measuring technological change essentially by comparing "maximal average product" in two different periods, which has little if any economic significance, leaves much to be desired.

An alternative bottoms-up approach is that of Peyrache [2014]. Following up on suggestions by Lovell [2003], Peyrache constructed a discrete approximation to the differential returns-to-scale coefficient evaluated at $\hat{y} = y/D_O(x, y, T)$ (a radial projection of y to the output possibility surface):

$$\epsilon(x, \hat{y}) = \left.\frac{\partial \ln D_O(\lambda x, \hat{y}, T)^{-1}}{\partial \ln \lambda}\right|_{\lambda=1}.$$

Distinguishing between an approximation from below and from above (owing to possible non-differentiability), Peyrache arrives at the following alternative indices measuring the contribution of returns to scale:

$$R(x^b, x^c, y^c, T^b) = \frac{D_O(x^b, \hat{y}^c, T^b) \ / \ D_O(x^c, \hat{y}^c, T^b)}{D_I(x^c, \hat{y}^c, T^b) \ / \ D_I(x^b, \hat{y}^c, T^b)}$$

and

$$R(x^b, x^c, y^b, T^b) = \frac{D_O(x^b, \hat{y}^b, T^b) \ / \ D_O(x^c, \hat{y}^b, T^b)}{D_I(x^c, \hat{y}^b, T^b) \ / \ D_I(x^b, \hat{y}^b, T^b)}.$$

In each case, the denominator is an input quantity index normalized on the frontier point $\hat{y}^b$ or $\hat{y}^c$. The numerator in each case is a measure of the radial expansion (or contraction if the value is less then 1) of the production possibility frontier along the ray through $\hat{y}^b$—equivalently, through y^b (owing to the first-degree homogeneity of D_O in y)—as the input vector changes from x^b to x^c.[27] Thus, in each case, the returns-to-scale index is a discrete measure of the aggregate increase in frontier output along a ray through the projected period-t frontier point, $\hat{y}^b$ or $\hat{y}^c$, as input quantities change between time b and time c.

Peyrache then defines his *radial productivity index*, Π_{RPI}, as the composition of the Malmquist output-oriented, technology-b-based productivity

[27] Bert Balk points out (in a private communication) that the numerators can also be characterized as dual input indices.

index (4.10)—which does not include scale effects—and the matching returns-to-scale index. Some manipulation, exploitating first-degree homogeneity of D_O in y, yields the following:

$$\begin{aligned}\Pi^c_{RPI}(x^b, x^c, y^b, y^c, T^b) &= \Pi^M_O(x^b, x^c, y^b, y^c, T^b) \cdot R(x^b, x^c, y^c, T^b) \\ &= \frac{D_O(x^b, y^c, T^b) \;/\; D_O(x^b, y^b, T^b)}{D_I(x^c, \hat{y}^c, T^b) \;/\; D_I(x^b, \hat{y}^c, T^b)} \end{aligned} \tag{4.29}$$

and

$$\begin{aligned}\Pi^b_{RPI}(x^b, x^c, y^b, y^c, T^b) &= \Pi^M_O(x^c, x^c, y^b, y^c, T^b) \cdot R(x^b, x^c, y^b, T^b) \\ &= \frac{D_O(x^c, y^c, T^b) \;/\; D_O(x^c, y^b, T^b)}{D_I(x^c, \hat{y}^b, T^b) \;/\; D_I(x^b, \hat{y}^b, T^b)}. \end{aligned} \tag{4.30}$$

The first index (4.29) normalizes on the current-period output ray, while the second (4.30) normalizes on the base-period output ray. As $\Pi^M_O(x^c, x^c, y^b, y^c, T^b)$ decomposes into a technological change component and an efficiency component, (4.29) and (4.30) yield tripartite decompositions of RPI productivity changes.

4.6 Non-Radial Indices and Indicators

The orientation of the Malmquist productivity index to either input space or output space poses a quandary—which orientation to use when. An approach to circumvention of this quandary is the (singular) measurement of productivity in the full ⟨input, output⟩ space (commonly referred to as "graph space"). Measurement in this space cannot follow a radial path to a (current- or base-period) frontier, since outputs must be expanded and inputs must be contracted. The most natural extension of the Malmquist index to this space is the hyperbolic index, a multiplicative expansion of output and contraction of input to a (current- or base-period) frontier of graph space. An alternative is the directional distance indicator, an *additive* expansion of output and contraction of input in a stipulated direction to a current- or base-period) frontier of graph space.[28] These functions are discussed in the next two subsections.

[28]Note that the Hicks-Moorsteen index does not measure productivity change along a stipulated path in graph space: rather, it is an amalgamation of separate measurements along radial paths to a frontier in input space and output space.

4.6.1 Hyperbolic Indices

As noted earlier, the distance-function components of Malmquist and Hicks-Moorsteen productivity indices are essentially equivalent to (simple transformations of) Debreu-Farrell technological efficiency indices. Extension of efficiency indices to ⟨input, output⟩ space is a precursor of extension of productivity indices to this space. This extension begins with the Färe, Grosskopf, and Lovell [1985] extension of the Debreu-Farrell efficiency index to ⟨input, output⟩ space:

$$E^H(x, y, T) = \min\{\lambda > 0 \mid \langle \lambda x, y/\lambda \rangle \in T\}.$$

As the path of the production vector to the frontier (as λ shrinks to its minimal level) is hyperbolic, Färe, Grosskopf, and Lovell refer to this index as the *hyperbolic efficiency index*. It maps (technologically feasible) production vectors and technologies into the $(0, 1]$ interval and is non-increasing in x and non-decreasing in y.

Analogously to the use of the Debreu-Farrell efficiency index to formulate the Malmquist index in input or output space, the hyperbolic efficiency index can be used to construct a "hyperbolic" productivity index in ⟨input, output⟩ space:

Technology b-based hyperbolic Malmquist index:

$$\Pi^H(x^b, x^c, y^b, y^c, T^b, T^c) = \frac{E^H(x^c, y^c, T^b)}{E^H(x^b, y^b, T^b)} \overset{\text{eff}}{=} E^H(x^c, y^c, T^b). \tag{4.31}$$

Technology c-based hyperbolic Malmquist index:

$$\Pi^H(x^b, x^c, y^b, y^c, T^b, T^c) = \frac{E^H(x^c, y^c, T^c)}{E^H(x^b, y^b, T^c)} \overset{\text{eff}}{=} \frac{1}{E^H(x^b, y^b, T^c)}. \tag{4.32}$$

Under the assumption of technologically efficient operation, this index measures productivity change as the "hyperbolic distance" from the current-period production vector to the base-period technology frontier or, alternatively, as the inverse of the hyperbolic distance from the base-period production vector to the current-period technological frontier. The index takes a value greater than one if and only if the technology frontier shifts upward in the hyperbolic direction in the neighborhood of the indicated production vector and hence is fundamentally a measure of technological change,

analogously to the Malmquist index under the assumption of technological efficiency. These indices satisfy the identity (I), monotonicity (M), and unit-invariance (UI) properties but fail to satisfy proportionality (P) or transitivity (T).

The middle terms in (4.31) and (4.32) reflect the possibility of technological inefficiency: *e.g.*, $E^H(x^b, y^b, T^b) < 1$ in (4.31) renders this term greater than the third term. Multiplication of top and bottom of these two terms by $E^H(x^c, y^c, T^c)$ and $E^H(x^b, y^b, T^b)$, respectively, yields decompositions of productivity change into indices of technical change and efficiency change:

$$\Pi^H(x^b, x^c, y^b, y^c, T^b, T^c) = \frac{E^H(x^c, y^c, T^b)}{E^H(x^c, y^c, T^c)} \; \frac{E^H(x^c, y^c, T^c)}{E^H(x^b, y^b, T^b)} \tag{4.33}$$

and

$$\Pi^H(x^b, x^c, y^b, y^c, T^b, T^c) = \frac{E^H(x^b, y^b, T^b)}{E^H(x^b, y^b, T^c)} \; \frac{E^H(x^c, y^c, T^c)}{E^H(x^b, y^b, T^b)}. \tag{4.34}$$

The first term on the right of each equation reflects technological change, measured in the hyperbolic direction and normalized alternatively on current-period and base-period production, while the second terms in each case reflects the change in efficiency The hyperbolic Malmquist index was proposed and implemented by Zofio and Lovell [2001].

4.6.2 Luenberger (Directional Distance) Indicators

The dominant non-radial formulation, however, has been the *Luenberger productivity indicator*, adapted from the shortage function of Luenberger [1992] to the measurement of productivity by Chambers [1996].[29] In the economics literature, Luenberger's concept has typically gone by the more evocative name, *directional distance function* (DDF). The DDF is defined by

$$\vec{D}(x, y, T, g) = \max\{\lambda \mid \langle x - \lambda g_x, \; y + \lambda g_y \rangle \in T\},$$

[29]See also Chambers, Chung, and Färe [1996], Chambers, Färe, and Grosskopf [1996], Färe and Grosskopf [1996], Balk [1998], and Chambers [1998, 2002]. A precursor of this literature is Diewert [1983], where a directional distance function is introduced to measure waste in production.

where $g = \langle g_x, g_y \rangle \in \mathbf{R}_+^{n+m}$. This function measures the feasible $\langle$contraction, expansion$\rangle$ of $\langle$input, output$\rangle$ quantities in the direction g and maps into $\mathbf{R}_+$.

Given our assumptions about the technology, the DDF is well-defined, non-decreasing in x, and non-increasing in y for all $\langle x, y, T \rangle \in \mathbf{N} \times \mathcal{T}$. As g lies in the same coordinate space as the production vector, the DDF is also independent of units of measurement. Moreover, $\vec{D}(x, y, T, g) \geq 0$ if and only if $\langle x, y \rangle \in T$, so that the DDF is a functional representation of the technology, and $\vec{D}(x, y, T, g) = 0$ if and only if $\langle x, y \rangle$ is contained in the frontier of T.

In contrast to the Debreu-Farrell-Malmquist distance function, which measures distance in ratio form, the DDF measures distance in terms of vector differences (as multiples of the direction vector g). Consequently, while the radial distance function has invariance properties with respect to certain rescaling of the data (owing to the homogeneity conditions), the DDF is invariant under transformations of the origin.

The *Luenberger productivity indicator* is defined as the arithmetic average of differences of directional distances from production vectors in periods b and c to a common technology (b or c):

$$\begin{aligned} \Pi_L(x^b, x^c, y^b, y^c, T^b, T^c) &= \frac{1}{2}\Big(\vec{D}(x^b, y^b, T^c, g) - \vec{D}(x^c, y^c, T^c, g) \\ &\quad + \vec{D}(x^b, y^b, T^b, g) - \vec{D}(x^c, y^c, T^b, g)\Big) \\ &=: \frac{1}{2}\Big(\bar{\Pi}_L^c(x^b, x^c, y^b, y^c, T^c) \\ &\quad + \bar{\Pi}_L^b(x^b, x^c, y^b, y^c, T^b)\Big); \end{aligned} \tag{4.35}$$

i.e., as the arithmetic average of a technology c-based Luenberger indicator and a technology b-based Luenberger indicator. Input-oriented and output-oriented Luenberger indicators are generated by setting $g_y = 0$ or $g_x = 0$.

The nomenclature *indicator*, as opposed to *index*, has been adopted to draw a distinction between difference-based measures like the Luenberger indicator and ratio-based measures like the Malmquist and Hicks-Moorsteen indexes. For theoretical comparisons of ratio-based and difference-based measures, see Chambers [1998, 2002] and Diewert [2005]. See Boussemart, Briec, Kerstens, and Poutineau [2003] for empirical comparisons of the two approaches to productivity measurement.

Analogously to the multiplicative decomposition of the Malmquist index (4.24)–(4.27), the Luenberger indicator (4.35) can be additively decomposed as follows:

$$\begin{aligned}\Pi_L(x^b, x^c, y^b, y^c, T^b, T^c) &= [\vec{D}(x^b, y^b, T^b, g) - \vec{D}(x^c, y^c, T^c, g)] \\ &+ \frac{1}{2}(\vec{D}(x^c, y^c, T^c, g) - \vec{D}(x^c, y^c, T^b, g) \\ &\quad + \vec{D}(x^b, y^b, T^c, g) - \vec{D}(x^b, y^b, T^b, g)) \quad (4.36)\end{aligned}$$

where the first term (in brackets) is efficiency change and the second term is technological change. This indicator satisfies properties analogous to those of the hyperbolic index.

Briec [1997] proposed a variation of the directional distance function by using the definition of $\vec{D}$ with the direction $g = \langle x, y\rangle$ and thereby specifying a specific distance measure rather than a class of measures parameterized by g. The *proportional directional distance function*, defined by

$$\vec{D}_P(x, y, T) = \max\Big\{\lambda \mid \langle(1-\lambda)x,\ (1+\lambda)y\rangle \in T\Big\},$$

maps into the [0, 1] interval. Replacing the directional distance function in (4.36) yields a *proportional Luenberger productivity indicator*. Note that this indicator, roughly speaking, is an additive analog of the hyperbolic index.

4.7 Dual Productivity Indices

The Farrell technological efficiency index lies at the core of the decomposition of productivity indices in Section 4.5. Farrell, however, was interested in a broader concept, encompassing *allocative* as well as technological efficiency. While the latter notion is a measure of the excess cost (above the minimum) attributable to operating below the production frontier (above the isoquant), the former is a measure of the extra cost accrued by operating at an inefficient point on the frontier, given input prices. Thus, examination of allocative efficiency, unlike technological efficiency, necessarily entails information about prices and economic units' adjustments to price changes. A thorough examination of economic efficiency would expand considerably the scope of this chapter beyond the common notion of productivity as a

technological phenomenon. Nevertheless, the recent introduction of notions of dual productivity indices suggests that a brief discussion is called for.

The cost function, $C : \mathbf{R}^n_{++} \times \mathbf{R}^m_+ \times \mathcal{T} \to \mathbf{R}_+$, dual to the input distance function, is defined by

$$\begin{aligned} C(p, y, T) &= \min_x \{p \cdot x \mid \langle x, y \rangle \in T\} \\ &= \min_x \{p \cdot x \mid D_I(x, y, T) \geq 1\}, \end{aligned}$$

where, of course, p is the input price vector. Under the CCD assumption of technological efficiency, the obvious candidate for a cost-based productivity index is simply the ratio of minimal costs in two situations normalized on particular output and price vectors: $C(y^t, p^t, T^b)/C(y^t, p^t, T^c)$, $t = b$ or c, and of course the geometric average of the two.

Assume the possibility of inefficient production. Cost efficiency is defined and decomposed into allocative and technological efficiency as follows:[30]

$$\begin{aligned} CE(p, x, y, T) := \frac{C(p, y, T)}{p \cdot x} &= \frac{C(p, y, T)}{p \cdot x / D_I(x, y, T)} \cdot \frac{1}{D_I(x, y, T)} \\ &=: AE_I(p, x, y, T) \cdot E_I(x, y, T). \end{aligned} \tag{4.37}$$

Recall that the Malmquist "ideal" (input-oriented) efficiency index (equation (4.15) can be written as

$$\Pi^{MI}_I(x^b, x^c, y^b, y^c, T^b, T^c) = \left(\frac{E_I(x^c, y^c, T^b)}{E_I(x^b, y^b, T^b)} \, \frac{E_I(x^c, y^c, T^c)}{E_I(x^b, y^b, T^c)} \right)^{1/2}. \tag{4.38}$$

The dual to this index, first formulated by Balk [1998, 67–70] (and recently re-examined by Zelenyuk [2006]), is the Malmquist ideal *cost-based productivity index*, a geometric average of cost-based productivity indices normalized on technologies b and c:

$$\begin{aligned} C\Pi^M(p^b, p^c, x^b, x^c, y^b, y^c, T^b, T^c) &= \left(\frac{CE(p^c, x^c, y^c, T^b)}{CE(p^b, x^b, y^b, T^b)} \, \frac{CE(p^c, x^c, y^c, T^c)}{CE(x^b, y^b, p^b, T^c)} \right)^{1/2} \\ &=: \left(C\Pi^M_b(p^b, p^c, x^b, x^c, y^b, y^c, T^b) \; C\Pi^M_c(p^b, p^c, x^b, x^c, y^b, y^c, T^c) \right)^{1/2} \end{aligned} \tag{4.39}$$

To obtain some intuition about this index, consider the case of no technological or allocative inefficiency in either period. In this case $CE(p^b, x^b, y^b, T^b) =$

[30] Note that $x/D_I(x, y, T) \in I(y)$ is the maximally contracted input vector.

1 and the cost-based productivity index normalized on technology b simplifies to

$$C\Pi_b^M(p^b, p^c, x^b, x^c, y^b, y^c, T^b) = CE(p^c, x^c, y^c, T^b) = \frac{C(p^c, y^c, T^b)}{p^c \cdot x^c}.$$

If $T^b \subset T^c$ and $\Gamma(b) \cap \Gamma(c) = \emptyset$ (reflecting global technological improvement), $CE(p^c, x^c, y^c, T^b) > 1$, indicating increased cost-based productivity (under the assumption of technological efficiency). Also, if there is allocative or technological inefficiency in period b, $CE(p^b, x^b, y^b, T^b) < 1$, providing additional fillip to the increase in measured productivity. The technology-c based index $C\Pi_c^M(p^c, p^b, x^b, x^c, y^b, y^c, T^c)$ can be similarly deconstructed.

Revenue-based "Malmquist" indices are analogously derived using the revenue function (dual to the output distance function),

$$R(r, x, T) = \max_y \{r \cdot y \mid \langle x, y \rangle \in T\} = \max_y \{r \cdot y \mid D_O(x, y, T) \leq 1\},$$

(where $r \in \mathbf{R}_{++}^m$ is the output price vector) and revenue efficiency indices analogous to the cost efficiency indices (4.37). See Balk [1998, 107–112] and Zelenyuk [2006] for details.

4.8 Aggregation of Productivity Indices

Productivity measurement is carried out at many levels of aggregation—*e.g.* at the firm level, the industry/sector level, and the economy-wide level. Some efforts have been made to establish conditions for consistent aggregation across these units of observation.

Index the individual economic units by $k = 1, \ldots, K$, so that the profiles of input vectors, output vectors, and technologies in period t are $\langle x_1^t, \ldots, x_K^t \rangle$, $\langle y_1^t, \ldots, y_K^t \rangle$, and $\langle T_1^t, \ldots, T_K^t \rangle$. The aggregate technology in period t is the Minkowski sum, $\bar{T}^t = \sum_k T_k^t$. Denote the aggregate production vector in period t by $\langle \bar{x}^t, \bar{y}^t \rangle := \langle \sum_k x_k^t, \sum_k y_k^t \rangle \in T^t$.

The strongest form of consistent aggregation, often referred to as "exact aggregation," is the existence of a productivity index Π and an aggregation rule Ξ such that

$$\Pi(\bar{x}^b, \bar{x}^c, \bar{y}^b, \bar{y}^c, \bar{T}^b, \bar{T}^c)$$

$$= \Xi\Big(\Pi(x_1^b, x_1^c, y_1^b, y_1^c, T_1^b, T_1^c), \ldots, \Pi(x_K^b, x_K^c, y_K^b, y_K^c, T_K^b, T_K^c)\Big). \quad (4.40)$$

To my knowledge, sufficient conditions (restrictions on the productivity index Π and the aggregation rule Ξ) for this strong form of aggregation have not been proved, and given related results in Blackorby and Russell [1999], tolerable sufficient conditions are highly unlikely to exist: they showed that, substituting efficiency indices for productivity indices, (4.40) holds only if production functions are highly linear (entailing linear isoquants and linear production possibility surfaces) and congruent across economic units.

Limited aggregation results have been established for dual productivity indices, owing to the well known (Koopmans [1957]) principle of interchangeability of set summation and optimization: *e.g.*,

$$C\left(p, y^1, \ldots, y^K, \sum_k T^k\right) = \sum_k C(p, y^k, T^k).$$

This fact, combined with (4.37), yields

$$CE\left(p, x^1, \ldots, x^K, y^1, \ldots, y^K, \sum_k T^k\right) = \sum_k CE(p, y^k, T^k) \cdot S^k(p, x^k), \quad (4.41)$$

where $S^k(p, x^k) = p \cdot x^k / p \cdot \sum_k x^k$ is the kth unit's share of aggregate cost. A suitable elaboration of (4.41) to encompass multiple periods and substitution into (4.39) yields a cost-based aggregate productivity index.

While this aggregation result has some power, since the aggregate index depends only on the aggregate technology, rather than the profile of technologies, it does not satisfy aggregation consistency, since aggregate cost efficiency, and hence aggregate productivity, depends on the entire distribution of both input and output quantities. Given the Koopmans interchangeability principle, this exercise yields little more than a simple summation: the aggregate economic unit really plays no specific role in the analysis.[31]

[31] Weaker aggregation conditions have been explored by Zelenyuk [2006] and Balk [2016].

4.9 Concluding Remarks (Malmquist vs. Hicks-Moorsteen Redux)

The Malmquist and Hicks-Moorsteen indices, developed respectively by Caves, Christensen, and Diewert [1982] and Bjurek [1996], persist to this day as the foundation of much of the pure theory of technology-based productivity indices. While the building blocks for both are Malmquist-Shephard distance functions, the two edifices are conceptually very different and give the same answer only under very restrictive conditions.

Consequently, there has been some discussion (and even a little controversy) in the literature about which is the appropriate measure—or even which is the one correct measure—of productivity change. Some of the discussion is almost semantic, revolving around the "appropriate" definition of "productivity". On the one hand, some make the valid point that the common understanding of productivity is "total-factor productivity"—an aggregate output quantity divided by aggregate input quantity—in which case a productivity index should be an output index divided by an input index.[32] The Hicks-Moorsteen index has this property and, unlike the Malmquist index, incorporates the effects of returns to scale (as well as input and output mix effects) into the measurement of productivity.

Consider, on the other hand, a comparison of the "productivity" of a firm in two periods b and c, and suppose that the production possibility surface of a firm at time c is above that at time b at all input quantities but, owing to diminishing returns to scale and a lower output in period c, total factor productivity at c is lower than that at b. It seems natural to say that the firm is more productive at time c, and that the lower total factor productivity at time c, when prices are different, is simply the result of optimizing behavior and hence does not reflect lower productivity of the firm. (If the technology at time b is not a proper subset of that at c, it is natural to compare the two in the neighborhood of production in period b or c, or at the mean of the

[32] This position has been most ardently argued by O'Donnell [2012], who goes on to develop an array of such ("multiplicatively complete") indices and proposed decompositions. In contrast to most of the theoretical literature on multiple-output productivity indices, however, O'Donnell's notion of total-factor productivity indices comprises "mechanistic" aggregate outputs and inputs—*i.e.,* aggregates that are not necessarily integral to the specification of the technology.

two.)

It seems to me that there is not a "correct" choice between the Malmquist and Hicks-Moorsteen indices: the choice may depend on the context of the productivity question (and the proclivities of the researcher), and the important thing is to be cognizant of the differences and take them into account in analyses, particularly in the interpretation of results.

From a purely axiomatic point of view, however, the Hicks-Moorsteen index would seem to have the upper hand, since it satisfies the proportionality condition (P) in addition to the conditions satisfied by the Malmquist index. (Both fail the transitivity test (T).[33])

While it is tempting to differentiate between the Malmquist index and the Hicks-Moorsteen index by renaming the former as an index of technological change, reserving the term productivity index for the latter, this would be misleading when we take into account the possibility of technologically inefficient production (production below the technological frontier): in this case the Malmquist index incorporates the effects of a change in technological efficiency as well as technological change (as does the Hicks-Moorsteen index).

While Malmquist indices decompose cleanly into technological and efficiency components, decomposition of the Hicks-Moorsteen index and further decomposition of the Malmquist index (into scale and mix effects) have been fraught with difficulties, though creative efforts have generated some interesting results.

Another nice attribute of Malmquist indices is that they can be extended quite naturally to measurement in the full ⟨input, output⟩ space. (The Hicks-Moorsteen index, by way of contrast, is a hybrid of separate measurements in input space and in output space.) The most natural extension is the hyperbolic index of technological change, a fairly straightforward generalization of the hyperbolic efficiency index. For reasons that escape me, this index has not flourished. In fact, several other technological efficiency indices in ⟨input, output⟩ space could be used to construct technology-based productivity indices.[34]

[33]But see footnote 22 above.

[34]See Russell and Schworm [2011] for descriptions and evaluations of these efficiency indices.

As noted earlier, the dominant non-radial productivity measure is based on the directional distance function. The DDF is actually a *class* of functions parameterized by the directional vector g. The discretion that this parameterization yields to the researcher is a double-edged sword. On the one hand, it opens up the possibility of incorporating a weighting scheme for inputs and outputs that reflects legitimate policy objectives. On the other hand, too much discretion can impose an unwanted burden on the researcher or, heaven forbid, can make productivity measurement subject to specious whims of the researcher. Empirical researchers have commonly disposed of this burden by setting the direction g equal to the unit vector. Although the DDF *formally* satisfies unit invariance (UI), when conjoined with this convention in the selection of g, it violates (UI) *in practice*.

Two other strands of research on theoretical productivity indices, dual indices and aggregation of indices, in my view have not yet led to important theoretical breakthroughs. Technological productivity is an important component of economic efficiency, but I think it is useful to maintain a distinction between purely internal productivity considerations and market oriented considerations. It is possible to derive quasi-aggregation results if some (optimal) reallocation of inputs or outputs among production units is incorporated into the analysis, but this approach really circumvents the aggregation problem by effectively converting the separate decison-making units into single decisions-making units.

To summarize, it seems to me that theoretical developments since the trailblazing CCD and Bjurek papers, while impressive, are not as important as the rich set of applications based on their indices. These applications are described in several chapters of this volume.

References

Balk, B. M. [1998], *Industrial Price, Quantity, and Productivity Indices: The Micro-Economic Theory and An Application*, Boston: Kluwer Academic Publishers.

Balk, B. M. [2001], "Scale Efficiency and Productivity Change," *Journal of Productivity Analysis* 15: 159–183.

Balk, B. M. [2016], "Various Approaches to the Aggregation of Economic

Productivity Indices," *Pacific Economic Review* 21: 445–463.

Balk, B. M., and R. Althin [1996], "A New, Transitive Productivity Index, *Journal of Productivity Analysis* 7: 19–27.

Berg, A., F. R. Førsund, and E. S. Jansen [1992], "Malmquist Indexes of Productivity Growth During the Deregulation of Norwegian Banking," *Scandinavian Journal of Economics* 94 (Supplement): S211–S228.

Bjurek, H. [1996], "The Malmquist Total Factor Productivity Index" *Scandinavian Journal of Economics* 98: 303–313.

Blackorby, C., and R. R. Russell [1999], "Aggregation of Efficiency Indexes," *Journal of Productivity Analysis* 12: 5–20.

Boussemart, J-P, W. Briec, K. Kerstens, and J-C Poutineau [2003], "Luenberger and Malmquist Productivity Indices: Theoretical Comparisions and Empirical Illustration," *Bulletin of Economic Research* 55: 391–405.

Briec, W. [1997], "A Graph-Type Extension of Farrell Technical Efficiency Measure," *Journal of Productivity Analysis* 8: 95–110.

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.

Chambers, R. G. [1996], "A New Look at Input, Output, Technical Change and Productivity Measurement," University Of Maryland Working Paper.

Chambers, R. G. [1998], "Input and Output Indicators," in Färe, Grosskopf, and Russell [1998], 241–271.

Chambers, R. G. [2002], "Exact Nonradial Input, Output, and Productivity Measurement," *Economic Theory* 20: 751–765.

Chambers, R. G., Y. Chung, and R. Färe [1996], "Benefit and Distance Functions," *Journal of Economic Theory* 70: 407–419.

Chambers, R. Färe, and Grosskopf [1996], "Productivity Growth in APEC Countries," *Pacific Economic Review* 1: 181–190.

Debreu, G. [1951], "The Coefficient of Resource Utilization," *Econometrica* 19: 273–292.

Diewert, W. E. [1983], "The Measurement of Waste within the Production Sector of an Open Economy," *Scandinavian Journal of Economics*, 85:

159–179.

Diewert, W. E. [1987], "Index Numbers," in J. Eatwell, M Milgate and P. Newman (eds.), *The New Palgrave. A Dictionary of Economics*, London: The Macmillan Press Limited: 767–780.

Diewert, W. E. [1992a], "Fisher Ideal Output, Input, and Productivity Indexes Revisited," *Journal of Productivity Analysis* 3: 211–248.

Diewert, W. E. [1992b], "The Measurement of Productivity," *Bulletin of Economic Research* 44: 163–198.

Diewert, W. E. [2005], "Index Number Theory Using Differences Rather Than Ratios," *American Journal of Economics and Sociology* 64: 311–360.

Färe, R., and S. Grosskopf [1996], *Intertemporal Production Frontiers: With Dynamic DEA*, Boston: Kluwer Academic Publishers.

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

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," *Economics Letters* 53: 269–274.

Färe, F. S. Grosskopf, and R. R. Russell (eds.) [1998], *Index Numbers: Essays in Honor of Sten Malmquist*, Boston: Kluwer Academic Publishers.

Färe, R., and D. Primont [1995], *Multi-Output Production and Duality: Theory and Applications*, Boston: Kluwer Academic Press.

Farrell, M. J. [1957], "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, 120: 253–290.

Grifell-Tatjé, E., and C. A. K. Lovell [1995], "A Note on the Malmquist Productivity Index," *Economics Letters* 47: 169–175.

Grifell-Tatjé, E., and C. A. K. Lovell [1999], "A Generalized Malmquist Productivity Index," *Sociedad de Estadística e Investigación Operativa TOP*, 7: 81–101.

Grosskopf, S. [2003], "Some Remarks on Productivity and Its Decomposi-

tions," *Journal of Productivity Analysis* 20: 459–474.

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

Koopmans, T. J. [1957], *Three Essays on the State of Economic Anaysis*, New York: McGraw-Hill.

Lovell, C. A. K. [2003], "The Decomposition of Malmquist Productivity Indexes," *Journal of Productivity Analysis* 20: 437–458.

Luenberger, D. G. [1992], "Benefit Functions and Duality," *Journal of Mathematical Economics* 21 : 115–145.

Malmquist, S. [1953], "Index Numbers and Indifference Curves," *Trabajos de Estatistica* 4: 209–242.

Moorsteen, R. H. [1961], "On Measuring Productive Potential and Relative Efficiency," *Quarterly Journal of Economics* 75: 451–467.

O'Donnell [2012], "An Aggregate Quantity Framework for Measuring and Decomposing Productivity Change," *Journal of Productivity Analysis* 38: 255–272.

Pastor, J. T., and C. A. K. Lovell [2005], "A Global Malmquist Productivity Index," *Economics Letters* 88: 266–271.

Peyrache, A. [2014], "Hicks–Moorsteen Versus Malmquist: A Connection by Means of a Radial Productivity Index," *Journal of Productivity Analysis* 41: 187–200.

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. [1998], "Distance Functions in Consumer and Producer Theory," in Färe, R., S. Grosskopf, and R. R. Russell [1998]: 7–90 .

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: Princeton University Press.

Solow, R. M. [1957], "Technical Change and the Aggregate Production Func-

tion," *Review of Economics and Statistics* 39: 312–320.

Wheelock, D. C., and P. W. Wilson [1999], "Technical Progress, Inefficiency, and Productivity Changes in U.S. Banking, 1984–1993," *Journal of Money, Credit, and Banking* 31: 212–234.

Zelenyuk, V. [2006], "Aggregation of Malmquist Productivity Indexes," *European Journal of Operational Research* 174: 1076–1086.

Zofio, J. L. [2007], "Malmquist Productivity Index Decompositions: a Unifying Framework," *Applied Economics* 39: 2371–2387.

Zofio, J. L., and C. A. K. Lovell [2001], "Graph Efficiency and Productivity Measures: an Application to US Agriculture," *Applied Economics* 33: 1433–1442.