INDEX NUMBERS: ESSAYS IN HONOUR OF STEN MALMQUIST

INDEX NUMBERS: ESSAYS IN HONOUR OF STEN MALMQUIST

EDITED BY

Rolf Färe

Southern Illinois University at Carbondale
Carbondale, Illinois 62901

Shawna Grosskopf

Southern Illinois University at Carbondale
Carbondale, Illinois 62901

R. Robert Russell

University of California at Riverside
Riverside, California

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data
Index numbers : essays in honor of Sten Malmquist / edited by Rolf Färe, Shawna Grosskopf, R. Robert Russell.
p. cm.
"Selections of papers presented at the Arne Ryde Symposium in honor of Professor Sten Malmquist, held in Lund on May 28, 1996" --Pref.
Includes bibliographical references and index.
ISBN 978-94-010-6035-6 ISBN 978-94-011-4858-0 (eBook)
DOI 10.1007/978-94-011-4858-0
1. Economics, Mathematical--Congresses. 2. Index numbers (Economics)--Congresses. I. Malmquist, Sten, 1917- . II. Färe, Rolf, 1942- . III. Grosskopf, Shawna. IV. Russell, R. Robert.
V. Arne Ryde Symposium (1996 : Lund, Sweden)
HB135.I547 1997
330'.01'51--dc21 97-38202
CIP

Originally published by Kluwer Academic Publishers in 1998
Softcover reprint of the hardcover 1st edition 1998

Printed on acid-free paper.

CONTENTS

PREFACE

This book contains a selection of papers presented at the Arne Ryde Symposium in honor of Professor Sten Malmquist, held in Lund on May 28, 1996. The symposium was financed by the Arne Ryde Foundation and organized by Bjørn Thalberg, Lund University; Rolf Färe, Southern Illinois University; and Pontus Roos, Institute of Health Economics, Lund.

The Arne Ryde Foundation was founded in memory of Arne Ryde, an exceptionally promising graduate student at the Department of Economics at the University of Lund. He died after an automobile accident in 1968 when he was only 23 years old. In his memory his parent established the foundation for the advancement of research at the Department of Economics at the University of Lund. The foundation finances international symposia in major fields of economic research.

The papers at the symposium were offspring of Professor Malmquist's work on index numbers. The symposium was honored by the presence of Professor Malmquist, now Professor Emeritus at the Department of Statistics at Stockholm University.

The manuscript has been prepared by Mariann Baratta; we thank her for a job well done.

INTRODUCTION

by

Rolf Färe, Shawna Grosskopf and R. Robert Russell

In 1952, Professor Sten Malmquist was appointed as an opponent on Dr. Erland von Hoffsten's doctoral dissertation "Price Indexes and Quality Changes," Uppsala Universitet. To facilitate his evaluation of this dissertation, Professor Malmquist constructed the now-famous Malmquist quantity index. In the course of formulating the index, Malmquist developed a distance function defined on a consumption space. This function, which is the consumer analog to the Shephard input distance function of producers, was used in ratio form to define the quantity index.

Out of Malmquist's opposition emerged the paper, "Index Numbers and Indifference Surfaces." The paper was subsequently presented at a statistical conference in Spain by Professor Hermand Wold, who at the time had no paper ready for the conference. In 1953, the paper was published in the Spanish statistical journal, *Trabajos de Estadistica* (Malmquist (1953)).

This volume contains papers based on Malmquist's contribution some 47 years ago. Some of the papers are surveys or expositions of the literature emanating from Malmquist's paper. Others are original contributions to the ongoing research agenda inspired by Malmquist's contribution.

Essay 1, by R. Robert Russell, is an exposition of the Malmquist-Shephard distance function and some of its uses in economic analysis. The essay describes the role of the distance function as an implicit representation of a (multiple output) technology (as well as a consumer preference ordering) and argues that it is a natural index of technical efficiency. The essay then develops the symmetric duality between the distance function and the (producer) cost function or (consumer) expenditure function. This essay also features the original insight of Malmquist's paper, in which the distance function is interpreted as a quantity index and a ratio of the distance function evaluated at different points constitutes an elegant index of the cost of living that is symetrically dual to the Konüs cost-of-living index. Finally, the essay also introduces the use of the distance function in formulating productivity indexes measuring changes in productivity over time or relative productivity levels of different production units. The emphasis of this exposition of the distance function and its applications is on intuition, emphasizing graphical illustrations, with references to the literature for proofs.

Essay 2, by Bert M. Balk, extends Malmquist's ideas to construct input-price, quantity, and productivity indexes for revenue-constrained producers—that is, firms that minimize cost subject to the attainment of a target revenue. The theory of revenue-constrained firms was developed by Färe and Grosskopf (1994) and has recently been used by Fisher (1995) to model a small open economy. In this essay, Balk lays out the basic duality theory that underlies the construction of his indexes and provides new results on the approximation of price and quantity indexes by Fisher index numbers and on the equivalence (under certain conditions) of his indexes and Törnqvist index numbers. His productivity indexes capture the effect of efficiency changes as well as technical change.

The third essay, jointly authored by Rolf Färe, Shawna Grosskopf, and Pontus Roos, is a survey of the theoretical and empirical work on Malmquist productivity indexes. On the theoretical side, the survey includes a number of issues that have arisen since the Malmquist productivity index was first proposed in the seminal paper by Caves, Christensen and Diewert (1982). These issues include the definition of the Malmquist productivity index; although all are based on the distance functions that Malmquist employed to formulate his original quantity index, variations include the geometric-mean form used by Färe, Grosskopf, Lindgren and Roos (1989) and the quantity index form referred to by Diewert (1992) as Hicks-Moorsteen. This raises the issue of whether this index is a technology index or an index of total-factor productivity (in the average-product sense), which in turn raises the issue of the role of returns to scale in specifying and computing the index. Other theoretical issues discussed here include the relationship to the Törnqvist and Fisher indexes, and the role of the circular test.

The survey of the empirical literature covers over 70 papers and is divided into sections reviewing studies of the public sector, banking, agriculture, countries and international comparisons, electric utilities, transportation, and insurance. Recommendations concerning the specification, computation, and interpretation of Malmquist productivity conclude this essay.

Essay 4 documents the ongoing implementation of Malmquist productivity indexes in the evaluation of the 800 member pharmacies in the Corporation, and is authored by two of the participants in that process, Nils-Olov Norlander (the Chief Controller of the Corporation) and Pontus Roos (the project manager at the Institute for Health Economics). This essay highlights the symbiotic relationship that has arisen during the process of this implementation: issues raised by individual pharmacies and the administration of the corporation became research topics in the development of the index itself, including the role of quality, the connection to profitability, *etc.* This relationship emphasizes the fact that this index is useful to industry as a performance tool precisely because it has been developed within a production-theoretic framework.

In Essay 5, a trio of Scandinavians, Hans Bjurek, Finn Førsund, and Lennart Hjalmarsson do an empirical comparison of various formulations of Malmquist productivity indexes. The basic two indexes are the geometric mean formulations of the Caves, Christensen and Diewert index and the Hicks-Moorsteen index, which they call the Malmquist Total Factor Productivity Index. These indexes are defined and discussed in Essay 3, and here they are empirically compared under different returns-to-scale assumptions. In total they calculate productivity for five different Malmquist indexes and the general consensus is that the different indexes track each other well.

In the last essay, Robert Chambers uses the directional technology distance function to characterize welfare indicators. Indicators are additive index numbers and the directional distance function is defined as a radial change in inputs and outputs in an arbitrary direction. Hence it generalizes the Malmquist-Shephard radial distance functions described in Essay 1. As shown by Chambers, the directional distance function has the additive structure required to characterize indicators like the Bennet-Bowley measures and the Luenberger (1995) input-output indicators.

References

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

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

Färe, R., and S. Grosskopf (1994) *Cost and Revenue Constrained Production*, Billant University Lecture Series Volume 4, Springer-Verlag, New York.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1989, 1994) "Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach," in A. Charnes, W.W. Cooper, A. Lewin and L.

Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston: Kluwer Academic Publishers.

Fisher, F. M. (1995) "The Production-Theoretic Measurement of Input Price and Quantity Indices," *Journal of Econometrics* 65, 155-174.

Luenberger, D. G. (1995) *Microeconomic Theory*, McGraw Hill, New York.

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

1

DISTANCE FUNCTIONS IN CONSUMER AND PRODUCER THEORY

by

R. Robert Russell*

1.1 INTRODUCTION

Although nobody seems to have recognized it at the time, 1953 was a banner year for duality theory, which in turn facilitated a veritable revolution in the application of consumer and producer theory some two decades later.[1] Sten Malmquist (1953) and Ronald W. Shephard (1953) independently introduced the notion of a distance function (also called a gauge function and a transformation function[2]) to economists.[3] In production theory, the (input) distance function

*I am grateful to Rolf Färe, Shawna Grosskopf, Craig Gundersen, Marc Mecurio, and especially Bert Balk for many insightful comments on an earlier draft. As I have not always followed their advice, they bear no blame for the remaining inadequacies.

[1]The two main innovations underlying this revolution were the application of duality theory and the formulation of flexible functional forms; Erwin Diewert (1971, 1974) led the charge in both areas.

[2]The appropriate nomenclature would seem to be "gauge function," since "distance function" has a different meaning in mathematics and "transformation function" has a traditional meaning in production theory that is more general than the Malmquist/Shephard distance function. But "distance function" is probably too firmly entrenched to dislodge.

[3]Debreu (1951) also formulated the distance function, which he called the "coefficient of resource utilization" (for reasons that will be explained later), but he did not develop the duality relationships that are a primary focus of this paper.

is simply the maximal radial contraction (equivalently, the minimal radial expansion) of an input vector consistent with the technological feasibility of producing a given output vector. In utility theory, it is the maximal radial contraction (or minimal radial expansion) of a consumption vector consistent with the attainment of a particular utility level.

Both contributions lay dormant until the 1970s, when they were rediscovered by the economics profession.[4] Shephard's application to production theory, especially as refined in his later volume (Shephard (1970)), finally found its way into the early-1970's circulation of manuscripts destined for the Fuss / McFadden (1978) volume (see especially McFadden (1978) and Hanoch (1978)). Malmquist's application to consumer theory was finally acknowledged in the survey of the theory of the cost-of-living index by Pollak (1971), but had been overlooked in earlier studies (*e.g.*, the classic survey of consumption theory by Houthakker (1961)).

As is now well known, the beauty and usefulness of the Malmquist/ Shephard distance function is partly attributable to its elegant (symmetric) duality with the more-familiar cost (or expenditure) function. Under reasonable regularity conditions, the properties of the distance function in (input or consumption) quantities are identical to the properties of the cost function in prices, and each is derived from the other by a simple, and essentially identical, algorithm. Moreover, the cost function can be alternatively characterized as a distance function in price space and the distance function can be alternatively characterized as an (imputed) cost function in price space. Finally, just as the gradient of the expenditure function is the Hicksian (vector valued) demand function (Shephard's Lemma), the gradient of the distance function is the inverse Hicksian demand function, and the Hessian of the distance function (the Antonelli matrix) has the same properties as the Slutsky matrix.

[4]Shephard's contribution was virtually ignored originally, perhaps because his book, prescient as it was at the time, contains little economic intuition and is a bit difficult to read. Malmquist's contribution was originally ignored because of the (literal) ignorance of the profession, perhaps because it was published in an obscure Spanish statistics journal.

Malmquist's contribution went beyond the introduction of the distance function to consumer theory: he formulated a quantity (or standard-of-living) index that is dual to the price (or cost-of-living) index originally introduced by Konüs (1924). In particular, the Malmquist quantity index is given by the ratio of values of the distance function (at different quantity situations, normalized on a given utility level), just as the Konüs price index is given by the ratio of values of the cost function (at different price situations, normalized on a given utility level). In short, the Malmquist quantity index and the Konüs price index are precisely dual to one another. With his focus on production theory, Shephard's contribution extended the notion of a distance function to output space—the maximal radial expansion (or minimal contraction) of an output vector consistent with the feasibility of producing the output vector with given input quantities. One purpose of this paper is to survey other applications of the distance functions, some of which were not anticipated by the progenitors.

What follows is much more an exposition than a survey. Hence, with apologies to the many researchers who have contributed substantially to the theory and use of distance functions, I make no attempt at exhaustive references to the literature. In addition, the emphasis is on intuition, making extensive use of graphical illustrations, with citations to the literature for proofs. In a sense, I would consider this a primer on the role of the distance function in economics. As such, it is unabashedly promotional.

One might question whether an exposition of distance functions and their applications is consistent with the "index number" theme of this volume. It is. As "index" simply means something general like "indicator," the operative word is *number*. The key feature of index numbers is that they aggregate a vector of variables into a scalar indicator (of the price level, the cost of living, the quantity level, the standard of living, productivity, the degree of (in)efficiency, *etc.*). That is the motif for this exposition of the Malmquist/Shephard distance function.

Section 1.2 develops the notions of the input and output distance functions. The focus of this section is on the distance functions as "natural" representations of multiple-output technologies; in the case of single-output technologies or preference theory, the production function and the utility function are perfectly servicable functional representations. Section 1.3 promotes the distance functions as appealing indexes of production efficiency (or inefficiency).

Section 1.4 develops the symmetric duality between the (input) distance function and the cost (expenditure) function and the exploitation of this duality for comparative-static demand analysis. Section 1.5 returns to the principal theme of Malmquist's classic paper, developing the theory of quantity indexes, or standard-of-living indices, as natural counterparts to the Könus price indexes, or cost-of-living indexes. Section 1.5 also applies the notion of Malmquist quantity indexes to the measurement of productivity changes or productivity differences among production units. Finally, Section 1.6 concludes.

1.2 DISTANCE-FUNCTION REPRESENTATIONS OF TECHNOLOGIES OR PREFERENCE ORDERINGS

Let $u \in \mathbf{R}_+^m$ and $x \in \mathbf{R}_+^n$ represent output and input quantity vectors, respectively. In most of the following, u and x can alternatively represent a utility scalar ($m = 1$) and a consumption vector, respectively. The primitive notion is that of a *technology set*,

$$T = \{\langle u, x\rangle \in \mathbf{R}_+^{m+n} \mid x \text{ can produce } u\}. \tag{1.2.1}$$

Alternatively, $\langle u, x\rangle \in T \subseteq \mathbf{R}_+^{1+n}$ if x generates utility at least equal to u. I assume throughout that T is *closed*.[5]

[5]For empirical applications, this assumption is inconsequential, since it cannot be refuted using any finite data base.

For all $u \in \mathbf{R}^m_+$, define the *input (consumption) requirement sets* by

$$L(u) = \{x \in \mathbf{R}^n_+ \mid \langle u, x\rangle \in T\} \tag{1.2.2}$$

and, for all $x \in \mathbf{R}^n_+$, define the *output possibility sets* by

$$P(x) = \{u \in \mathbf{R}^m_+ \mid \langle u, x\rangle \in T\}. \tag{1.2.3}$$

Thus, $L(u)$ is the set of input vectors that can produce u (alternatively, the set of consumption vectors that generate utility at least equal to u) and $P(x)$ is the set of output vectors that can be produced with input vector x. Clearly, $L(u)$ is closed for all u, $P(x)$ is closed for all x, and

$$\langle u, x\rangle \in T \iff x \in L(u) \iff u \in P(x), \tag{1.2.4}$$

so that the technology (or preference ordering) is completely characterized by the correspondence, $L : \mathbf{R}^m_+ \mapsto \mathbf{R}^n_+$, or the correspondence, $P : \mathbf{R}^n_+ \mapsto \mathbf{R}^m_+$, as well as by the set T (the graph of L and of P). Figure 1.1 illustrates output-possibility and input-requirement sets.

Define the u-isoquant (or u-indifference-curve) by

$$I(u) = \{x \in L(u) \mid \lambda x \notin L(u) \; \forall \; \lambda < 1\}. \tag{1.2.5}$$

Thus, the u^*-isoquant (indifference curve) in Figure 1.1A is the boundary of $L(u^*)$ relative to $\mathbf{R}^n_+$.[6] Note that, for $m > 1$, these isoquants can intersect. Similarly, the u-production-possibility-curve is defined by

$$\Gamma(u) = \{u \in P(x) \mid \lambda u \notin P(x) \; \forall \; \lambda > 1\}, \tag{1.2.8}$$

and is given by the boundary of $P(x^*)$ relative to $\mathbf{R}^m_+$ in Figure 1.1B.

---

[6]Define the ϵ-neighborhood of x relative to $\mathbf{R}^n_+$ by

$$N^+_\epsilon(x) = \left\{\hat{x} \in \mathbf{R}^n_+ \;\middle|\; \|\hat{x} - x\| < \epsilon\right\}, \tag{1.2.6}$$

where $\|x\|$ is the Euclidean norm of x. The boundary of $L(u^*)$ relative to $\mathbf{R}^n_+$ is

$$\left\{x \in \mathbf{R}^n_+ \mid N^+_\epsilon(x) \cap L(u) \neq \emptyset \;\wedge\; N^+_\epsilon(x) \cap \left(\mathbf{R}^n_+ \backslash L(u)\right) \neq \emptyset \; \forall \; \epsilon > 0\right\}. \tag{1.2.7}$$

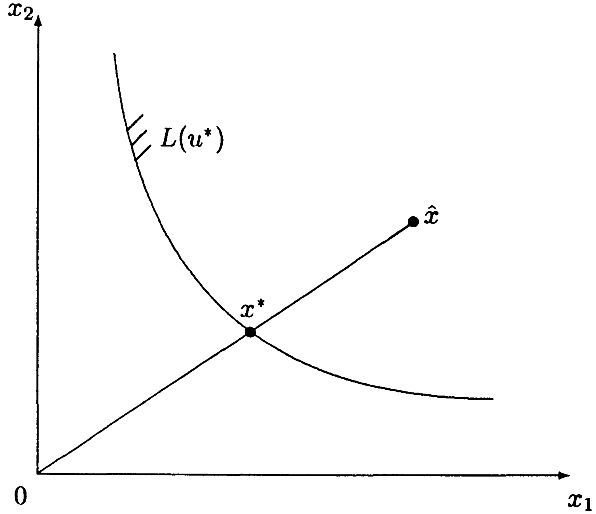

Figure 1.1A

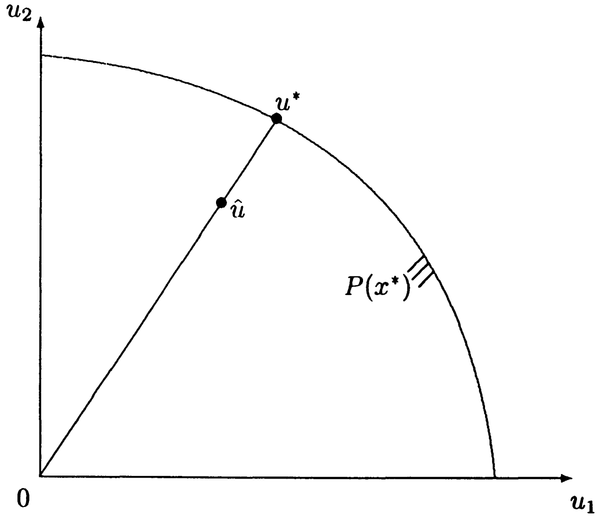

Figure 1.1B

For many purposes, and for most applications, it is useful to characterize the technology or preference ordering by a function rather than a set (or set-valued mapping). In the case of utility theory or a single-output production technology, a functional representation is straightforward. For $m = 1$, define the production (utility) function by

$$U(x) = \max\{u \mid \langle u, x\rangle \in T\}. \tag{1.2.9}$$

Then, assuming free disposability of the output,

$$\langle u, x\rangle \in T \iff u \leq U(x), \tag{1.2.10}$$

or, for $U(x) > 0$,[7]

$$\langle u, x\rangle \in T \iff \frac{u}{U(x)} =: F(u, x) \leq 1. \tag{1.2.11}$$

The implicit representation F of the production function U is useful because it not only indicates whether the output/input combination is (output) efficient—whether $F(u, x) = 1$, but also constitutes an index of the amount of inefficiency: the ratio of actual to potential output.

In the case of multiple-output production technologies, the approach to constructing a functional representation of the technology is less obvious. The traditional textbook approach is to adopt the same scalar, extreme-value approach as in the case of a single output. For example, define the function, $f^j : \mathbf{R}_+^{m-1+n} \to \mathbf{R}_+$, by

$$f^j\Big(u^{(-j)},\ x\Big) = \max\Big\{u_j \mid \langle u, x\rangle \in T\Big\}, \tag{1.2.12}$$

where the $m-1$ dimensional vector $u^{(-j)}$ is the output vector u with the j^{th} component purged. Then, if u_j is freely disposable,

$$\langle u, x\rangle \in T \iff u_j \leq f^j\Big(u^{(-j)}, x\Big), \tag{1.2.13}$$

and, for $f^j\Big(u^{(-j)},\ x\Big) > 0$,

[7] Notation: $A =: B$ or $B := A$ means that B is *defined* by A.

$$\langle u,x\rangle \in T \iff \frac{u_j}{f^j\left(u^{(-j)},\ x\right)} =: F^j(u,x) \leq 1. \tag{1.2.14}$$

Alternatively, define $g^i : \mathbf{R}_+^{m+n-1} \to \mathbf{R}_+$, by

$$g^i\left(u,\ x^{(-i)}\right) = \min\left\{x_i \mid \langle u,x\rangle \in T\right\}. \tag{1.2.15}$$

Then, if x_i is freely disposable,

$$\langle u,x\rangle \in T \iff x_i \geq g^i\left(u,\ x^{(-i)}\right), \tag{1.2.16}$$

and, for $g^i\left(u,\ x^{(-i)}\right) > 0$,

$$\langle u,x\rangle \in T \iff \frac{x_i}{g^i\left(u,\ x^{(-i)}\right)} =: G^i(u,x) \geq 1. \tag{1.2.17}$$

Either of these approaches characterizes the technology set, but there is an inelegant asymmetry about them. For example, $F^j(u,x) = 1$ if $\langle u,x\rangle$ is "efficient" in the coordinate direction of the j^{th} output and $G^i(u,x) = 1$ if $\langle u,x\rangle$ is "efficient" in the coordinate direction of the i^{th} input. But $\langle u,x\rangle$ could be inefficient in every other coordinate direction. Similarly, if $F^j(u,x) < 1$ or $G^i(u,x) > 1$, the values indicate the degree of inefficiency in terms of the proportional shortfall in the j^{th} output or the excessive utilization of the i^{th} input.[8] But these measures of inefficiency are arbitrary. How does one choose the coordinate direction in which to measure inefficiency? Why is one output or one input more indicative than any other? If one wanted the characterization to be independent of the normalization on a single input or a single output—that is, if one wanted a rule for representing a technology by a function without knowing ahead of time what the technology looked like, it would be necessary to posit a strong disposability condition: either[9]

[8]Some textbook treatments define, say, G^i by $G^i(u,x) = x_i - g^i\left(u, x^{(-i)}\right)$, so that $\langle u,x\rangle \in T \iff G^i(u,x) \geq 0$. While this function characterizes the production technology (assuming that the i^{th} input is freely disposable), it has no meaning for inefficient output/input combinations, since it is sensitive to the choice of units of measurement of the i^{th} input.

[9]Notation: $x \geq x'$ if $x_i \geq x'_i\ \forall i$, $x > x'$ if $x \geq x'$ and $x \neq x'$, and $x \gg x'$ if $x_i > x'_i\ \forall\, i$.

strong input disposability: $x \in L(u)$ and $x' \geq x$ implies $x' \in L(u)$ for all $u \in \mathbf{R}_+^m$ (equivalently, $x' \geq x$ implies $P(x) \subseteq P(x')$); or

strong output disposability: $u \in P(x)$ and $u' \leq u$ implies $u' \in P(x)$ for all $x \in \mathbf{R}_+^n$ (equivalently, $u' \leq u$ implies $L(u) \subseteq L(u')$).[10]

(In consumer theory, strong input disposability will be referred to as *strong nonsatiation.*)

In both panels of Figure 1.2, the required strong disposability assumption is violated. In Figure 1.2A, $G^1(u^*, \hat{x}) = b/a > 1$ but $\hat{x} \notin L(u^*)$ and hence $\langle u^*, \hat{x}\rangle \notin T$. In Figure 1.2B, $F^1(\hat{u}, x^*) = a/b < 1$ but $\hat{u} \notin P(x^*)$ and hence $\langle \hat{u}, x^*\rangle \notin T$.

An alternative approach that is symmetric (with respect to inputs or, alternatively, outputs) and that requires a weaker disposability assumption than does the above approach (for arbitrary coordinate normalizations) is to characterize the input-requirement set or the output possibility set (and hence the technology) by the maximal *equi-proportionate* contraction of *all* inputs or the maximal *equi-proportionate* expansion of *all* outputs consistent with keeping the output/input vector in the technology set. These are verbal descriptions of the input and output distance functions.

1.2.1 Input Distance Function

The *input distance function*, $D_I : \mathbf{N} \to \mathbf{R}_+$, is defined by

$$D_I(u, x) = \max\Big\{\lambda > 0 \mid x/\lambda \in L(u)\Big\}, \tag{1.2.18}$$

where[11]

$$\mathbf{N} = \Big\{\langle u, x\rangle \in \mathbf{R}_+^{m+n} \mid 0^{(n)} \notin L(u) \wedge \ x/\lambda \in L(u) \text{ for some } \lambda > 0\Big\}. \tag{1.2.19}$$

[10] Strong input disposability is also equivalent to $L(u) = L(u) + \mathbf{R}_+^n$ for all $u \in \mathbf{R}_+^m$ and strong output disposability is also equivalent to $(P(x) - \mathbf{R}_+^m) \cap \mathbf{R}_+^m \subseteq P(x)$ for all $x \in \mathbf{R}_+^n$.

[11] $0^{(n)}$ is the n-dimensional zero vector.

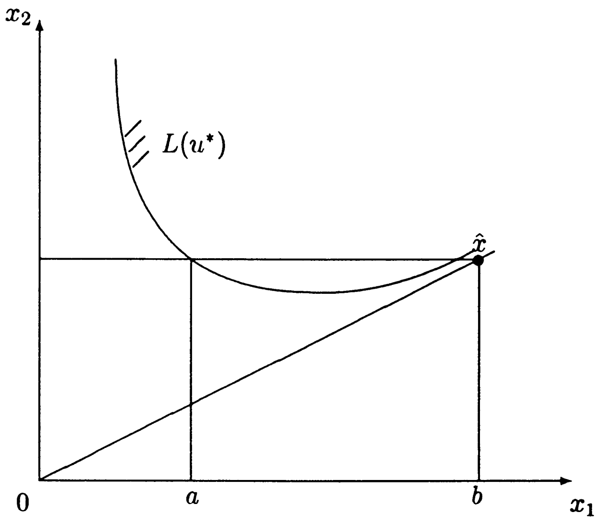

Figure 1.2A

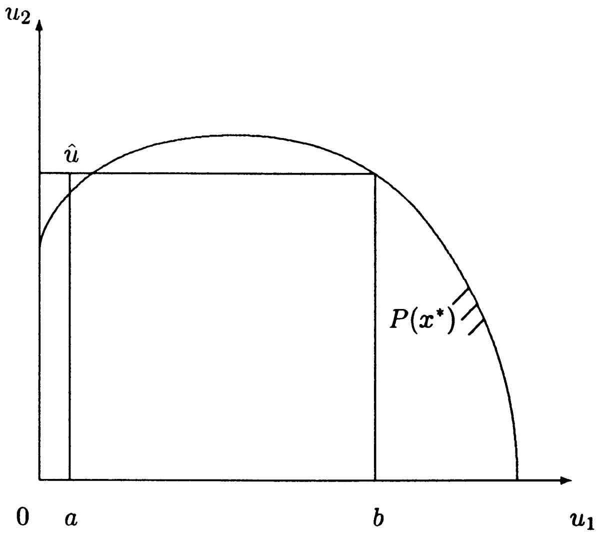

Figure 1.2B

Thus, in Figure 1.1A above, $D_I(u^*, \hat{x}) = \|\hat{x}\| \;/\; \|x^*\|$.

The need for the last domain restriction in defining D_I is apparent from Figure 1.2A, where there does not exist a $\lambda > 0$ such that $\hat{x}/\lambda \in L(u^*)$; the same problem exists if $x = 0^{(n)} \notin L(u)$. In addition, in order for the maximal λ to exist, any output vector (presumably $0^{(m)}$) that can be produced with no inputs must be purged from output space.[12] Closedness of $L(u)$ now guarantees that D_I is well defined on **N**.

(1.2.20) **Theorem**: D_I is homogeneous of degree one in x.[13]

This property is apparent from Figure 1.1A, but the general proof is also immediate, since, for any scalar $\theta > 0$,

---

[12]One could extend the domain to the entire $(m+n)$-dimensional Euclidean space, with the conventions that

(i) $D_I(u, 0^{(n)}) = 1$ if $0^{(n)} \in L(u)$,

(ii) $D_I(u, x) = +\infty$ if $0^{(n)} \in L(u)$ and $x \neq 0^{(n)}$,

and

(iii) $D_I(u, x) = 0$ if there exists no $\lambda > 0$ such that $x/\lambda \in L(u)$,

(with the corresponding extension of the range of D_I to include $+\infty$ in (ii)). What follows would remain true (with the usual arithmetic involving $+\infty$). An alternative approach is to define D_I as a supremum, $D_I(u, x) = \sup\{\lambda > 0 \mid x/\lambda \in L(u)\}$, to avoid these domain restrictions, but I believe it's a little more intuitive to restrict the domain sufficiently to avoid the need for this approach (and perhaps incorporate the three extensions above). Also, taking the supremum fails to deal adequately with the case (i) problem. If $x = 0^{(n)} \in L(u)$, then $\{\lambda > 0 \mid x/\lambda \in L(u)\} = (0, +\infty]$, and both the supremum and the maximum equal $+\infty$; but $0^{(n)} \in I(u)$, so that Theorem 1.2.25 below does not hold. (The supremum is also equal to $+\infty$ in case (ii), compatibly with the above intuition. In case (iii), $\{\lambda > 0 \mid x/\lambda \in L(u)\} = \emptyset$ and hence, vacuously, *any* $r \in \mathbf{R}_+ \cup \{+\infty\}$ is an upper bound on this set; hence, the least upper bound is zero, compatibly with the intuition of this case.)

[13]Note that the projection of the domain, **N**, onto input space (for any output vector) is a cone, since $x/\lambda \in L(u)$ for some $\lambda > 0$ implies that, for arbitrary $\theta > 0$, $\theta x/\lambda \in L(u)$ for some $\lambda > 0$; hence, $D_I(u, \theta x)$ is defined if $D_I(u, x)$ is. Note also that this result and those that follow hold without restricting the output/input vector to **N** with the extensions of footnote 12 (alternatively—except for Theorem 1.2.25—by defining D_I as the sup; see Färe and Primont (1995)).

$$\begin{aligned} D(u,\theta x) &= \max\{\lambda > 0 \mid \theta x/\lambda \in L(u)\} \qquad (1.2.21) \\ &= \theta \max\left\{\lambda/\theta > 0 \;\middle|\; \frac{x}{\lambda/\theta} \in L(u)\right\} \\ &= \theta D(u,x). \end{aligned}$$

Homogeneity of the input distance function, which is not satisfied by the asymmetric functional representations of technologies described on pages 13-14, will prove to be useful.

To guarantee that the input distance function characterizes the technology, a weak disposability assumption is needed:

Weak input disposability: $x \in L(u) \Rightarrow \lambda x \in L(u)\ \forall \lambda > 1$.

It is clear that this assumption is weaker than strong input disposability. For example, the input requirement set $L(u^*)$ in Figure 1.2A satisfies weak but not strong input disposability. (In the case of consumer theory, weak input disposability will be referred to as *weak nonsatiation*.)

(1.2.22) **Theorem**: For all $\langle u, x\rangle \in \mathbf{N}$, weak input disposability is satisfied if and only if

$$x \in L(u) \iff D_I(u,x) \geq 1, \qquad (1.2.23)$$

so that $L(u)$ is recovered from D_I by

$$L(u) = \{x \in \mathbf{R}^n_+ \mid D_I(u,x) \geq 1\}. \qquad (1.2.24)$$

The necessity of weak disposability is illustrated in Figure 1.3, where $D_I(u^*, \hat{x}) > 1$ but $\hat{x} \notin L(u^*)$.

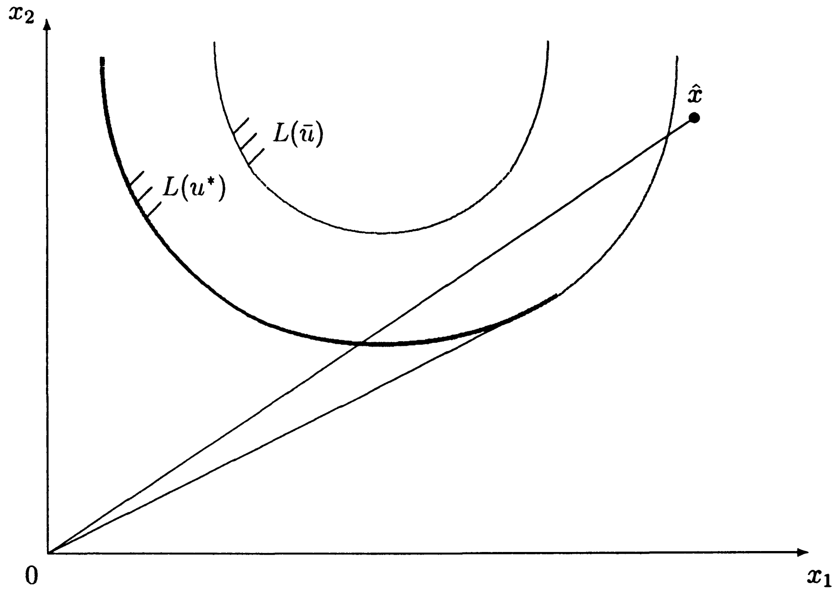

Figure 1.3

It is also possible to completely characterize the isoquant (or indifference curve) using the input distance function:

(1.2.25) **Theorem**: For all $\langle u, x\rangle \in \mathbf{N}$, $x \in I(u)$ if and only if $D_I(u, x) = 1$.[14]

Note that this characterization of the u-isoquant does not require the weak disposability assumption; in Figure 1.3, $I(u^*)$ is the darkened portion of the boundary of $L(u^*)$, which is characterized by $\{x \in \mathbf{R}^n_+ \mid D_I(u^*, x) = 1\}$.

[14]Again, the domain restriction can be eliminated by adopting (i)–(iii) in footnote 12. Note, however, that the sup approach would not work, because of the problem with case (i) pointed out in footnote 12.

Suppose that, in addition to strong output disposability, we posit that $I(\hat{u}) \cap I(u) = \emptyset$ if $\hat{u} \neq u$. In the case of single-output production or utility theory, this is equivalent to the intermediate-theory assumption that higher isoquants (indifference surfaces) represent greater output (utility) and isoquants (indifference surfaces) have no points in common. (The level sets in Figure 1.4, with the "pole" at $\hat{x}$ and $\bar{u} > u^*$, violate this latter condition.) If, in addition, L satisfies local non-saturation (or local non-satiation),[15] then

$$D_I(u,x) = 1 \iff u = U(x), \tag{1.2.27}$$

and U is obtained by straightforward inversion of $D(u,x) = 1$ in u (*i.e.*, $u = D_I^{-1}(1,x)$).

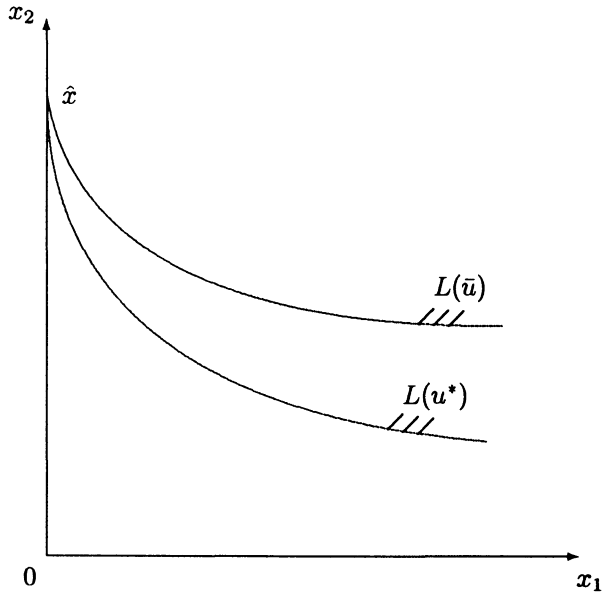

Figure 1.4

[15]Local non-saturation: for all $\langle u, x\rangle \in \mathbf{N}$ and for all $\epsilon > 0$, there exists an

$$\bar{x} \in N_\epsilon(x) = \{\hat{x} \in \mathbf{R}^n_+ \mid \|\hat{x} - x\| < \epsilon\} \tag{1.2.26}$$

such that $U(\bar{x}) > U(x)$.

The foregoing indicates that the input distance function constitutes a functional representation of the technology under very weak disposability assumptions. Moreover, this representation is homogeneous of degree one in input quantities for all technologies, including non-homothetic ones. Of course, additional properties of the distance function are implied by additional restrictions on the technology.

Input (or preference) convexity: $L(u)$ is convex for all $u \in \mathbf{N}$.

Continuity of L: The correspondence, $L : \mathbf{R}^m_+ \longmapsto \mathbf{R}^n_+$, is continuous (upper and lower hemi-continuous).

(1.2.28) **Theorem**: If the technology (or preference ordering) is closed and satisfies strong (input and output) disposability, input (preference) convexity, and continuity of L, D_I is (i) nonincreasing in u; (ii) nondecreasing, homogeneous of degree one, and concave in x; and (iii) jointly continuous in u and x for $x \gg 0^{(n)}$.

Property (i) follows immediately from strong output disposability. Monotonicity and concavity in x follow from strong input disposability and input convexity, while the homogeneity condition follows directly from the definition of D_I.[16] Joint continuity in u and x follows from strong disposability and the continuity of the correspondence L. Strong input disposability eliminates spikes, like that in Figure 1.5A. In this case, x^ν converges to $\hat{x}$ and $D_I(u^*, x^\nu)$ converges to $\|\hat{x}\| \;/\; \|x'\|$, which is less than $D(u^*, \hat{x}) = \|\;\hat{x}\|\;/\|x^*\|$. An example of a discontinuity of L is the case of a "thick" isoquant (indifference curve), as shown in Figure 1.5B. In this case, L is not lower hemi-continuous, since $\hat{x} \in L(u^*)$ and $u^\nu = 2^{1/\nu}u^*$ converges

[16]Proofs of these properties can be found in early duality works, like Samuelson (1960), Shephard (1970), Diewert (1974), Blackorby, Primont, and Russell (1978), Hanoch (1978), and McFadden (1978).

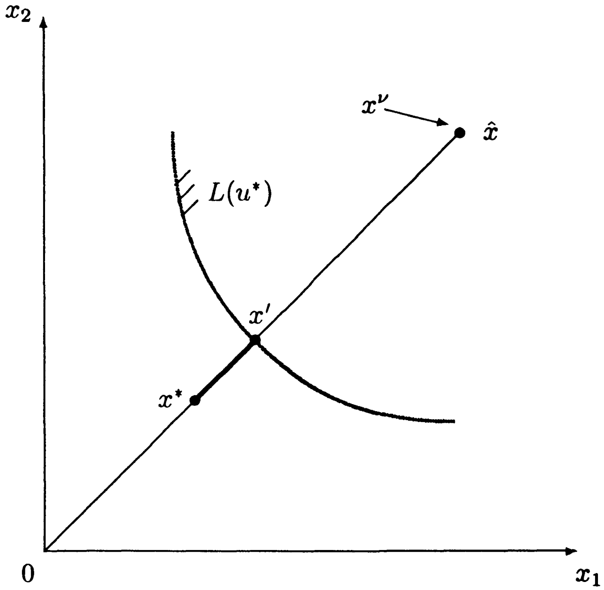

Figure 1.5A

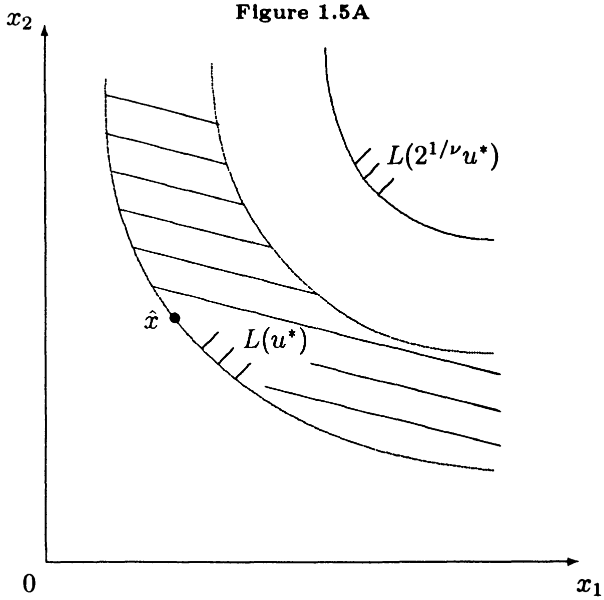

Figure 1.5B

to u^*, but there does not exist a sequence $\{x^\nu\} \in L(u^\nu)$ converging to $\hat{x}$. As a consequence, $D(u^\nu, \hat{x})$ converges to a value greater than $D(u^*, \hat{x}) = 1$.[17]

(1.2.29) **Corollary**: Given a D_I that satisfies (i), (ii), and (iii) in Theorem 1.2.28, there exists a technology (preference ordering), T, satisfying strong (input and output) disposability, input (preference) convexity, and continuity of L, such that D_I is derived from T by (1.2.18).

Thus, if one specifies an input distance function satisfying the appropriate monotonicity, homogeneity, concavity, and continuity properties, one knows that this distance function can be derived by (1.2.18) from a closed technology set T (or preference ordering) satisfying the strong disposability, input (preference) convexity, and continuity conditions.

1.2.2 Output Distance Function

A production technology can also be represented by an output distance function defined on the production possibility set in the same manner as the input distance function is defined on the input requirement set. One assumption is needed:

Boundedness: $P(x)$ is bounded for all $x \in \mathbf{R}^n_+$.

This assumption, along with closedness of T, guarantees that the production possibility set is compact.

[17]See Hackman and Russell (1995) for proof of the continuity property. A violation of upper hemi-continuity of L would be generated, for example, by a jump in outputs that can be produced with an infinitesimally small increase in some input. For the counterexample (provided by Rolf Färe) that forces us to restrict x to the interior of $\mathbf{R}^n_+$, see Hackman and Russell (1995).

The *output distance function*, $D_O : \mathbf{M} \longrightarrow \mathbf{R}_+$, is defined by

$$D_O(u,x) = \min\left\{\lambda > 0 \mid u/\lambda \in P(x)\right\}, \tag{1.2.30}$$

where

$$\mathbf{M} = \{\langle u,x\rangle \in \mathbf{R}_+^{m+n} \mid u \neq 0^{(n)} \ \wedge \ u/\lambda \in P(x) \text{ for some } \lambda > 0\}. \tag{1.2.31}$$

Thus, in Figure 1.1B, $D_O(\hat{u},x^*) = \|\hat{u}\| \,/\, \|u^*\|$. The output distance function is a natural generalization of the ratio of actual to potential output in the single-output case described on page 13. While there is no ambiguity about the direction in which to measure this ratio in the case of a scalar output, a direction must be chosen in the case of multiple outputs. The output distance function measures the output expansion in a radial direction and hence is symmetric in the outputs. The compactness of the production possibility set, along with the restriction of the domain to $\mathbf{M}$, guarantees that D_O is well defined. Figure 1.6 illustrates the need for the second domain restriction in (1.2.31): $D_O(\hat{u},x^*)$ is undefined.[18]

(1.2.32) **Theorem**: D_O is homogeneous of degree one in u.

This property is evident from the construction in Figure 1.1B, and the proof is essentially equivalent to the proof of homogeneity of the input distance function in x.[19]

For the output distance function to be an implicit representation of the technology, a disposability assumption, analogous to weak input disposability, is needed:

Weak output disposability: $u \in P(x)$ and $0 \leq \lambda \leq 1$ implies $\lambda u \in P(x)$.

[18]Rather than purging the zero vector from output space, one could define D_O by taking the infimum rather than the minimum: if $u = 0^{(m)} \in P(x)$, $\{\lambda > 0 \mid u/\lambda \in P(x)\}$ is equal to the interval $(0,+\infty)$, in which case the minumum does not exist, but the infimum equals zero. If there exists no λ such that $u/\lambda \in P(x)$, $\{\lambda > 0 \mid u/\lambda \in P(x)\} = \emptyset$, so that the infimum is $+\infty$.

[19]Again, as $\mathbf{M}$ is cone, $D_O(\theta u,x)$ is defined for $\theta > 0$ if $D_O(u,x)$ is.

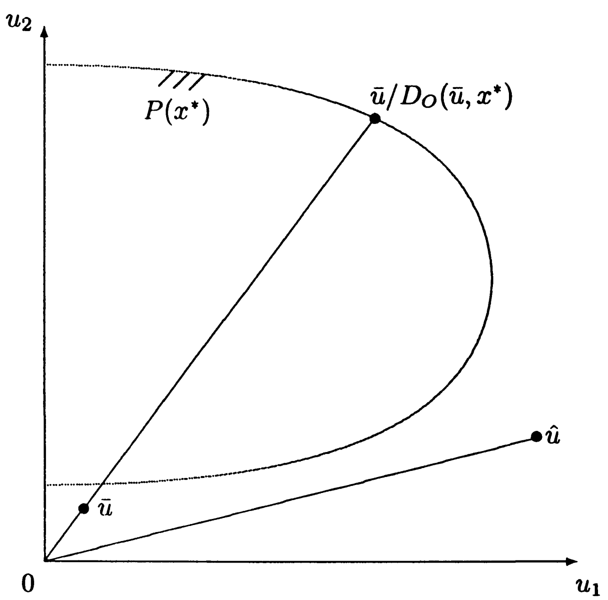

Figure 1.6

(1.2.33) **Theorem**: Weak output disposability is satisfied if and only if

$$u \in P(x) \iff D_O(u,x) \leq 1, \tag{1.2.34}$$

so that $P(x)$ is recovered by

$$P(x) = \{u \in \mathbf{R}^m_+ \mid D_O(u,x) \leq 1\}. \tag{1.2.35}$$

The need for weak disposability is illustrated in Figure 1.6, where $D(x^*,\bar{u}) < 1$ but $\bar{u} \notin P(x^*)$. This restriction is not needed for the following characterization of the production possibility curve:

(1.2.36) **Theorem**: For all $\langle u, x\rangle \in \mathbf{M}$, $u \in \Gamma(x)$ if and only if $D_O(u, x) = 1$.

Additional assumptions yield additional properties of the output distance function.

Output convexity: $P(x)$ is convex for all $x \in \mathbf{R}^m_+$.

Continuity of P: The correspondence, $P : \mathbf{R}^n_+ \mapsto \mathbf{R}^m_+$, is continuous (upper and lower hemi-continuous).

(1.2.37) **Theorem**: If the technology is closed and satisfies strong input and output disposability, output convexity, and continuity of the correspondence P, D_O is (i) nonincreasing in x; (ii) nondecreasing, homogeneous of degree one, and convex in u; and (iii) jointly continuous in u and x for all $\langle u, x\rangle$ such that $u \gg 0^{(m)}$.[20]

(1.2.38) **Corollary**: Given a D_O that satisfies (i), (ii), and (iii) in Theorem 1.2.37, there exists a technology, T, satisfying strong (input and output) disposability, output convexity, and continuity of P, such that D_O is derived from T by (1.2.30).

Thus, if one specifies an output distance function satisfying the appropriate monotonicity, homogeneity, convexity, and continuity properties, one knows that it can be derived from a closed technology set T satisfying the output disposability, input convexity, and continuity conditions.

Finally, the input and output distance functions are symmetrically dual, since they can be derived from one another by simple algo-

[20] Again, proofs can be found in Shephard (1970), Diewert [1974], and Hackman and Russell (1995). Discontinuity problems arise if there exist "thick" production possibility curves, in which case P is not lower hemi-continuous.

rithms:

$$\begin{aligned} D_O(u, x) &= \min\{\lambda > 0 \mid u/\lambda \in P(x)\} \\ &= \min\{\lambda > 0 \mid x \in L(u/\lambda)\} \\ &= \min\{\lambda > 0 \mid D_I(u/\lambda, x) \geq 1\} \end{aligned} \tag{1.2.39}$$

and

$$\begin{aligned} D_I(u, x) &= \max\{\lambda > 0 \mid x/\lambda \in L(u)\} \\ &= \max\{\lambda > 0 \mid u \in P(x/\lambda)\} \\ &= \max\{\lambda > 0 \mid D_O(u, x/\lambda) \leq 1\}. \end{aligned} \tag{1.2.40}$$

To summarize, under very weak disposability conditions, the input and output distance functions serve as alternative implicit representations of multiple-output technologies (and of single-output technologies and preferences as special cases). They are in some sense more natural than the traditional textbook constructions of implicit representations, since they are symmetric with respect to inputs or outputs.[21] In addition, unlike traditional representations, the distance functions are necessarily homogeneous in inputs or outputs, a property that is useful in some contexts. Additional properties of the distance functions follow from imposing some regularity conditions on the technology. Finally, as we shall see in the next section, the value of the index is meaningful when output/input combinations are inefficient.

[21] Each is, of course, arbitrary in its characterization in either input space or output space; generalizations of the distance function to the complete output/input space, referred to as "shortage functions" by Luenberger (1992, 1995) and as "directional distance functions" by Chambers (1996) and Chambers, Chung, and Färe (1996), are reviewed in Chambers' contribution to this volume. The first use of this concept was in the context of welfare economics in Blackorby and Donaldson (1980a).

1.3 DISTANCE FUNCTIONS AS TECHNICAL EFFICIENCY INDEXES

It appears that the (input) distance function actually made its first appearance in economics as a "coefficient of resource utilization" in Debreu (1951). As noted in the previous section, $D_I(u,x) > 1$ means that $x \in L(u)$ but $x \notin I(u)$, so that the input vector x can produce u, given the technology T, but the same output vector could be produced with less of all inputs if the production unit operated more efficiently. In other words, x is technically inefficient in producing u, in the sense of Koopmans (1951). Moreover, for all $\langle u,x\rangle \in T$, $D_I(u,x)^{-1}$ is an index of the degree of inefficiency, varying in value over the half-open interval $(0,1]$ and equal to 1 if (but not only if) x is efficient in the production of u. This notion of the measurement of inefficiency was applied to data on U.S. farms, using a (non-parametric) mathematical-programming method, by M. J. Farrell (1957). Thus, the *Debreu/Farrell input-based efficiency index*, $E_{DF} : \overset{\circ}{T} \to (0,1]$, is defined by

$$E_{DF}(u,x) = \frac{1}{D_I(u,x)}, \tag{1.3.1}$$

where

$$\overset{\circ}{T} = \{\langle u,x\rangle \in T \mid 0^{(n)} \notin L(u)\}. \tag{1.3.2}$$

Farrell's contribution was itself ignored for more than two decades. It was rediscovered by Charnes, Cooper, and Rhodes (1978), who referred to the mathematical-programming method of measuring technical efficiency as data envelopment analysis (DEA), an appellation that seems to have stuck.[22] Their paper has led to a flood of papers applying DEA, most of them in management-science/ operations-research journals.[23]

[22]See Färe, Grosskopf, and Lovell (1994) for a comprehensive development of the mathematical-programming approach to the measurement of technical efficiency.

[23]See Seiford (1996) for an extensive bibliography.

At the same time, Färe and Lovell (1978) noted that the Debreu/Farrell index of technical efficiency had some unattractive properties if technologies did not satisfy strong input disposability. Thus, in Figure 1.7, $1/D_I(u^*, \hat{x}) = 1$ although $\hat{x}$ is not efficient in the production of u^*. Also, although $2\hat{x} > 2x^*$, and hence $2\hat{x}$ is, in an obvious sense, less efficient in the production of u^*, nevertheless $D_I(u^*, 2\hat{x}) = D_I(u^*, 2x^*)$, so that $2\hat{x}$ and $2x^*$ receive the same efficiency score according to the Debreu/Farrell criterion.

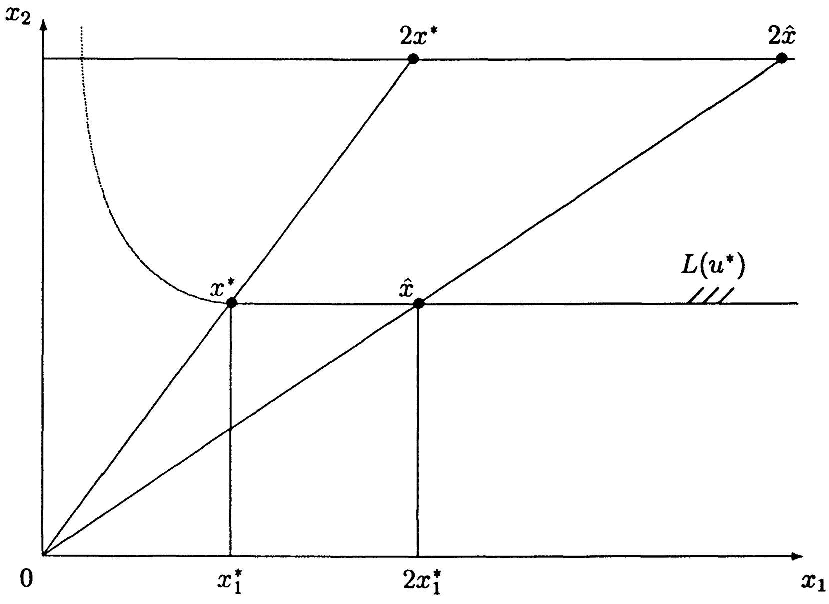

Figure 1.7

1.3.1 The Färe/Lovell Axioms

Färe and Lovell went on to suggest a set of three desirable axioms for (input based) efficiency indexes, $E : \mathring{T} \to (0, 1]$:[24]

Indication (I): $\forall \ \langle u, x\rangle \in \mathring{T}$, $E(u, x) = 1 \iff x$ is efficient in producing u.

Monotonicity (M): $\forall\langle u, x\rangle \in \mathring{T}$, $\hat{x} > x \Rightarrow E(u, \hat{x}) < E(u, x)$.

Homogeneity (H): $\forall\langle u, x\rangle \in \mathring{T}$, $E(u, \lambda x) = \lambda^{-1}E(u, x) \ \forall\lambda > 0$.

The Debreu/Farrell index clearly satisfies (H), but not (I) or (M), for all technologies.

Färe and Lovell went on to suggest an alternative efficiency index that measures inefficiency as the maximal sum of *coordinate-wise* (as opposed to radial), proportional contractions of inputs. They mischievously called this index the "Russell measure," but I call it the *Färe/Lovell efficiency index*, $E_{FL} : \mathring{T} \to (0, 1]$, defined by

$$E_{FL}(u, x) = \min_{\kappa}\left\{\frac{\sum_i \kappa_i}{\sum_i \delta(x_i)} \,\middle|\, \langle \kappa_1 x_1, \ldots, \kappa_n x_n\rangle \in L(u) \right.$$
$$\left. \wedge\ \kappa_i \in [0, 1] \forall\ i\right\},$$

where

$$\delta(x_i) = \begin{cases} 1 \text{ if } x_i > 0 \\ 0 \text{ if } x_i = 0 \end{cases}.$$

The motivation for this formulation is obvious: in Figure 1.7,

$$E_{FL}(u^*, \hat{x}) = (1/2 + 1)/2 = 3/4, \tag{1.3.3}$$

[24]They suggested a fourth as well, but Russell (1985) argued that this condition was not well defined and when it was given a rigorous definition it was implied by the other three axioms.

since $\hat{x}$ can be contracted to the efficient point x^*, and

$$\begin{aligned} E_{DF}(u^*, 2x^*) &= (1/2 + 1/2)/2 = 1/2 > \\ E_{FL}(u^*, 2\hat{x}) &= (1/4 + 1/2)/2 = 3/8. \end{aligned}$$

Unfortunately, the Färe/Lovell index was later shown to violate the homogeneity condition for some technologies, leading Zieschang (1984) to formulate an index that combines the Debreu/Farrell and Färe/Lovell index in a clever way to guarantee that the homogeneity condition is satisfied, as is the case for the Debreu/Farrell index, and that contracts any input vector to an efficient point, as is true of the Färe/Lovell index. Zieschang's concept is defined on the *free-disposal hull* of the input requirement set,

$$L^+(u) = L(u) + \mathbf{R}^n_+. \tag{1.3.4}$$

Denote the distance function defined on the free-disposable hull by

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

This augmentation of $L(u)$, along with the modified distance function, is illustrated in Figure 1.8.

One can now define the *extended Debreu/Farrell and Färe/Lovell efficiency indexes* on the free-disposal hulls:

$$E^+_{DF}(u,x) = \frac{1}{D_I^+(u,x)} \tag{1.3.5}$$

and

$$\begin{aligned} E^+_{FL}(u,x) &= \min_{\kappa}\left\{\frac{\sum_i \kappa_i}{\sum_i \delta(x_i)} \mid \langle \kappa_1 x_1, \ldots, \kappa_n x_n\rangle \in L^+(u)\right. \\ &\qquad \left. \wedge\ \kappa_i \in [0,1]\ \forall\, i\right\}. \end{aligned}$$

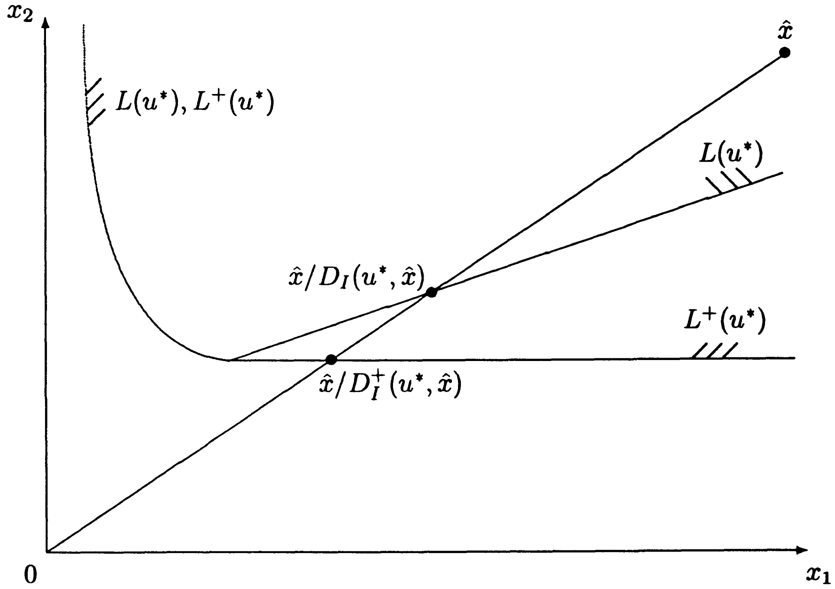

Figure 1.8

The *Zieschang efficiency index*, $E_Z(u, x) : \overset{\circ}{T} \longrightarrow (0, 1]$, is now defined by

$$\begin{aligned} E_Z(u, x) &= E_{FL}^+(u,\ E_{DF}^+(u, x)\ x) \ \times\ E_{DF}^+(u, x) \\ &= \frac{E_{FL}^+\Big(u,\ x/D_I^+(u, x)\Big)}{D_I^+(u, x)}. \end{aligned}$$

Thus, the Zieschang index first contracts input vectors radially to the isoquant of the free-disposal hull of the input requirement set and then in coordinate-wise directions to an efficient point. It is clearly homogeneous of degree -1 in input quantities, since D_I^+ is homogeneous of degree 1.

Unfortunately, this is not the end of the story. Russell (1985) showed that neither the Färe/Lovell nor the Zieschang index satisfies the monotonicity axiom for all closed technologies. These problems are illustrated in Figure 1.9.

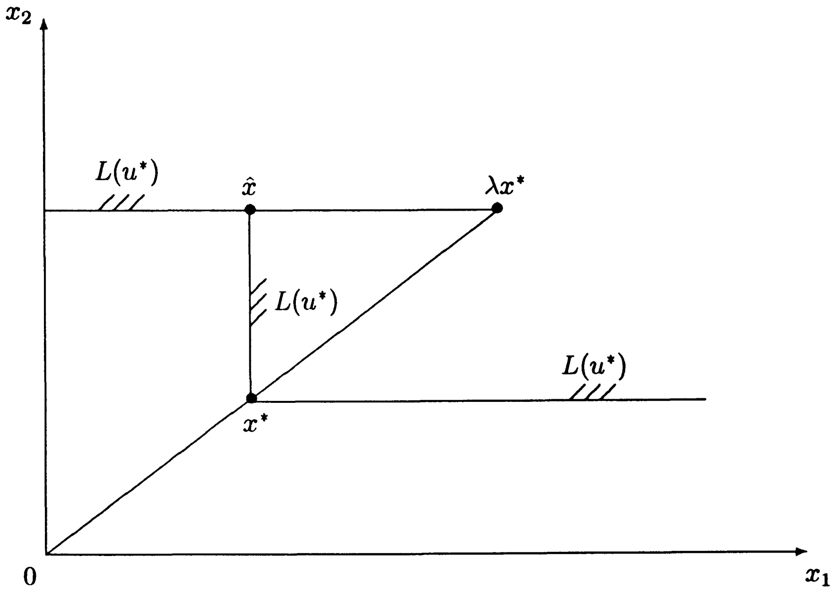

Figure 1.9

Suppose that $\lambda \in (1,2)$, in which case $\lambda^{-1} \in (1/2, 1)$. Then

$$E_{FL}(u^*, \lambda x^*) = \min\left\{\frac{\lambda^{-1} + \lambda^{-1}}{2}, \frac{0+1}{2}\right\} = \frac{1}{2} < \lambda^{-1},$$

indicating that E_{FL} violates (H), and

$$E_{FL}(u^*, \hat{x}) = \min\left\{\frac{1 + \lambda^{-1}}{2}, \frac{0+1}{2}\right\} = \frac{1}{2}, \tag{1.3.6}$$

showing that E_{FL} also violates (M). Now suppose that $\lambda = 2$, in which case

$$E_Z(u, \hat{x}) = \min\left\{\frac{1 + \frac{1}{2}}{2}, \frac{0+1}{2}\right\} = \frac{1}{2} = E_Z(u, \lambda x^*), \tag{1.3.7}$$

indicating that E_Z violates (M).

Shortly after the above problems were pointed out, the search for an efficiency index that satisfied the Färe/Lovell axioms for all technologies was brought to an abrupt halt by the impossibility theorem of Bol (1986):

(1.3.8) **Theorem**: There exists no efficiency measure satisfying (I), (M), and (H) for all technologies.

This result is easy to illustrate in two-space. The indication condition (I) implies that $E(u^*, x^*) < 1$ in Figure 1.10. By homogeneity (H), there exists a $\lambda > 1$ such that $E(u^*, \hat{x}) \in N_\epsilon(1)$ for arbitrary $\epsilon > 0$. These two facts imply that the monotonicity condition (M) is violated.[25]

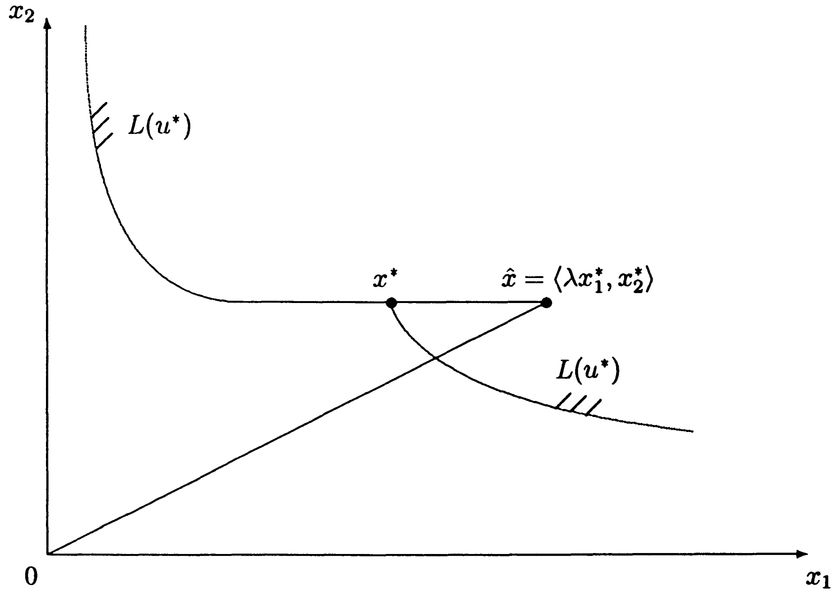

Figure 1.10

Bol identified two ways out of the conumdrum posed by his impos-

[25]While this, and many other two-space examples contain non-convex level sets, similar convex examples can be constructed in higher-order spaces. (See Bol (1986).)

sibility theorem: (1) to weaken the Färe/Lovell axioms or (2) to restrict the class of technologies to which the efficiency index is to be applied. I will briefly address each of these approaches.

1.3.2 Weakening the Färe/Lovell Axioms

An obvious candidate for weakening is monotonicity, since Russell (1987) showed that this axiom is incompatible with the fundamental requirement that an index be independent of the units of measurement. This property has been called "commensurability" by Eichhorn and Voeller (1976) (in the context of price indxees) and is defined as follows:

Commensurability (C): If $u^* = \Omega u$ and $x^* = Kx$, where Ω and K are positive diagonal matrices, and $L^*(u^*) = \{x^* \mid K^{-1}x^* = x \in L(u)\}$, then $E(u^*, x^*) = E(u, x)$.

(1.3.9) **Theorem**: There exists no efficiency measure satisfying (C) and (M) for all technologies.

This result is easy to illustrate in two-space. In Figure 1.11, $L(\hat{u}) = \{\bar{x}\} + \mathbf{R}^n_+$. (Note that this is not a particularly bizarre technology. It could correspond to the case where input 1 has zero marginal product until the amount of input 2 reaches the critical level $\bar{x}_2$ and has positive marginal product for $x_2 > \bar{x}_2$. The isoquants could be of quite conventional shapes for $x_2 > \bar{x}_2$.) Now consider point $\hat{x}$ on the $\hat{u}$-isoquant and suppose that the unit of measurement of input 1 only is changed; *i.e.*, Ω in the definition of commensurability is the identity matrix and

$$K = \begin{pmatrix} \kappa_1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.3.10}$$

where $\kappa_1 > 1$. Clearly, $L^*(u^*) =: L^*(\Omega\hat{u}) = L(\hat{u})$. Commensurability now implies that

$$E(\Omega\hat{u}, K\hat{x}) =: E(u^*, x^*) = E(u, x^*) = E(u, \hat{x}). \tag{1.3.11}$$

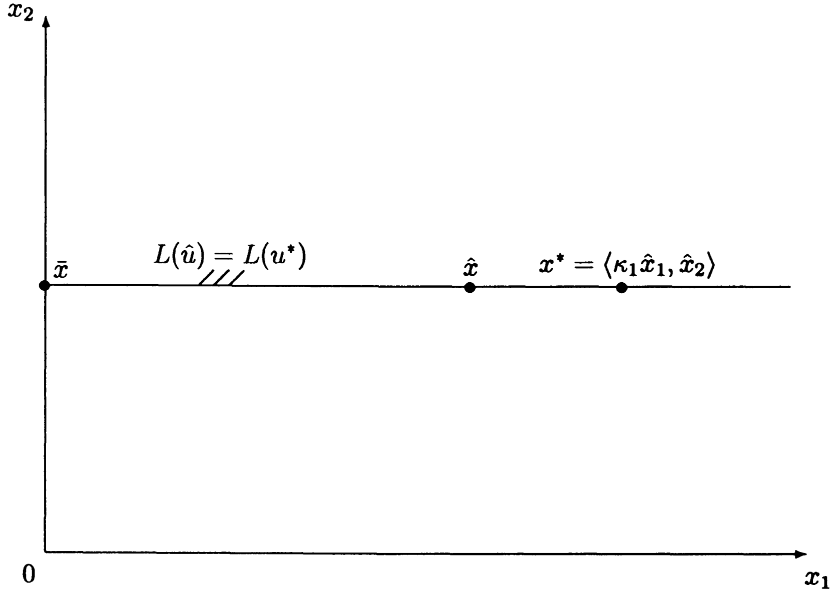

Figure 1.11

But this is inconsistent with (M), since $x^* > \hat{x}$. The key feature in this example is that the efficiency index can't tell the difference between a change in the units of input 1 (stretching the horizontal axis) and an increase in x_1.

It is therefore instructive to weaken the monotonicity condition to the following:

Weak monotonicity (WM): $\hat{x} \geq x \Rightarrow E(u, \hat{x}) \leq E(u, x)$.

In addition, since no radial efficiency index can satisfy the indication condition (I), let us consider an alternative indication condition. Let us write $x \stackrel{*}{>} \hat{x}$ if $x_i > \hat{x}_i$ for all i such that $x_i > 0$ and say that $x \in L(u)$ is *weakly efficient* if $x \stackrel{*}{>} \hat{x}$, $x \in \overset{\circ}{T}$, and $\hat{x} \in \overset{\circ}{T}$, implies $\hat{x} \notin L(u)$. Thus, points like $\hat{x}$ in Figure 1.7 are weakly efficient (but not Koopmans (1951) efficient), while points like $2\hat{x}$ are not weakly efficient. The $\overset{*}{>}$ concept is needed to accommodate situations like that in Figure 1.12, where $\bar{x}$ is not weakly efficient in $L(\bar{u})$, yet $\bar{x} \gg \tilde{x}$ and $\tilde{x} \in \overset{\circ}{T}$ vacuously implies $\tilde{x} \notin L(\bar{u})$, since there does not exist an $\tilde{x}$ satisfying $\bar{x} \gg \tilde{x}$ and $\hat{x} \in L(\bar{u})$. But $\bar{x} \overset{*}{>} \hat{x}$ and $\hat{x} \in L(\bar{u})$, so that $\bar{x}$ is not weakly efficient.

Indication of weakly efficient vectors (IW): $E(u, x) = 1 \iff x$ is weakly efficient.

This condition is neither weaker nor stronger than (I), since neither implies the other; nevertheless, it is convenient to refer to (I) as the stricter indication condition.

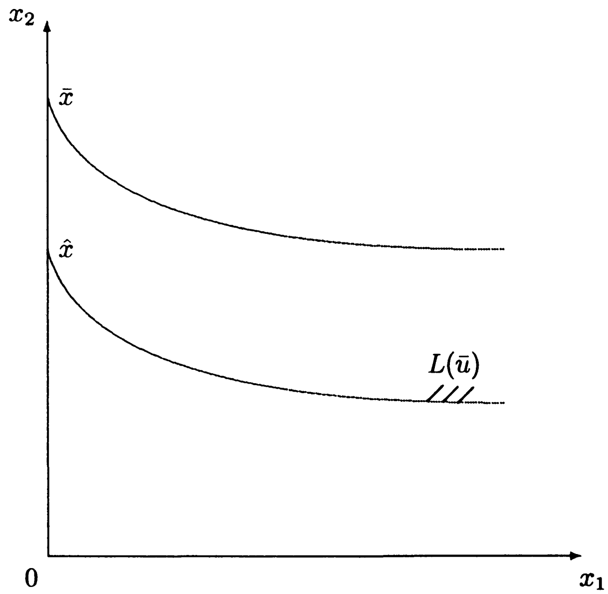

Figure 1.12

The following theorem summarizes the possibility results in the literature:[26]

[26]See Färe and Lovell (1978); Färe, Lovell, and Zieschang (1983); and Russell (1987).

(1.3.12) **Theorem**:

- E_{FL} satisfies (C), (I), and (WM).
- E_Z satisfies (C), (I), and (H).
- E^+_{DF} satisfies (C), (IW), (WM) and (H).

These results summarize the trade-offs between the three proposed efficiency indexes. One trades off

(i) weak monotonicity against homogeneity in comparing the Färe/Lovell and Zieschang indexes;

(ii) the stricter indication condition condition against homogeneity when comparing the Färe/Lovell and extended Debreu/Farrell indexes; and

(iii) the stricter indication condition against weak monotonicity in comparing the Zieschang and extended Debreu/Farrell indexes.

As homogeneity enjoys a venerable position in the axiomatic treatment of index numbers,[27] these trade-offs would seem to argue in favor of the extended Debreu/Farrell index, which, of course, is essentially the same as the Malmquist/Shephard distance function (restricted to feasible output/input vectors and extended to the free-disposal hull of the input requirement set). The next subsection provides an additional argument in favor of the distance function in measuring productive efficiency.

1.3.3 Continuity Axioms

Russell (1990) suggested that continuity is another desirable axiom for efficiency indexes, especially where measurement error exists, since it provides some assurance that "small" measurement errors of

[27] See Eichhorn and Voeller (1976) and Balk (1995) for reviews of this literature.

input and output quantities result only in "small" errors of efficiency measurement. This axiom turns out to interact in a substantive way with one of the Färe/Lovell axioms and has implications for the choice of a rule for measuring efficiency.

(1.3.13) **Theorem**: There does not exist an efficiency measure that satisfies (I) and continuity in either u or x for all technologies.

This result is easily demonstrated diagrammatically.[28] Assume that (I) holds and consider the convergent sequences $u^\nu \to u^o$ in Figure 1.13A and $x^\nu \to x^o$ in Figure 1.13B. In Figure 1.13A, $E(u^\nu, \bar{x}) = 1$ for all ν, but $E(u^o, \bar{x}) < 1$.[29] In Figure 1.13B, $E(\bar{u}, x^\nu) = 1$ for all ν, but $E(\bar{u}, x^o) < 1$. Of course either a discontinuity in u for fixed x or a discontinuity in x for fixed u implies that E fails to satisfy joint continuity in $\langle u, x \rangle$.

As both the Färe/Lovell and the Zieschang measures satisfy (I), Theorem 1.3.13 says that they cannot satisfy any continuity condition for all (closed) technologies. On the other hand, an immediate implication of Theorem 1.2.28 is

(1.3.14) **Theorem**: If the technology satisfies continuity of L, the extended Debreu/Farrell efficiency index, E^+_{DF}, is jointly continuous in u and x for $x \gg 0^{(n)}$.

Note that the counterexamples in Figure 1.13 are consistent with the assumptions in Theorem 1.3.13. Thus, the introduction of continuity axioms provides additional justification for the use of the distance function (defined on the free-disposal hull of the input requirement set) as an index of technological efficiency.

[28] For a proof, see Russell (1990), where it is also shown that monotonicity is incompatible with continuity, maintaining the (IW) condition.

[29] This convergence of $L(u^\nu)$ to $L(u^o)$ could be generated by a technology in which the marginal product of input 2 in producing, say, output 1, is zero, $u_2^\nu \to 0$, and u_1^ν is calibrated so that $\bar{x}$ is on the u^ν-isoquant for all ν.

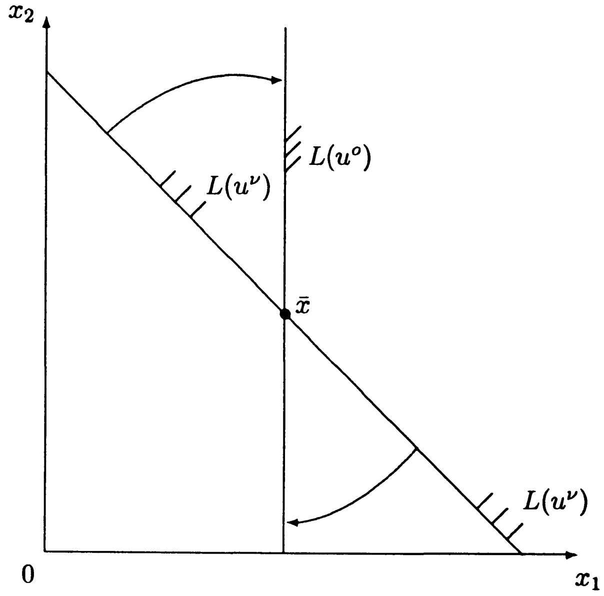

Figure 1.13A

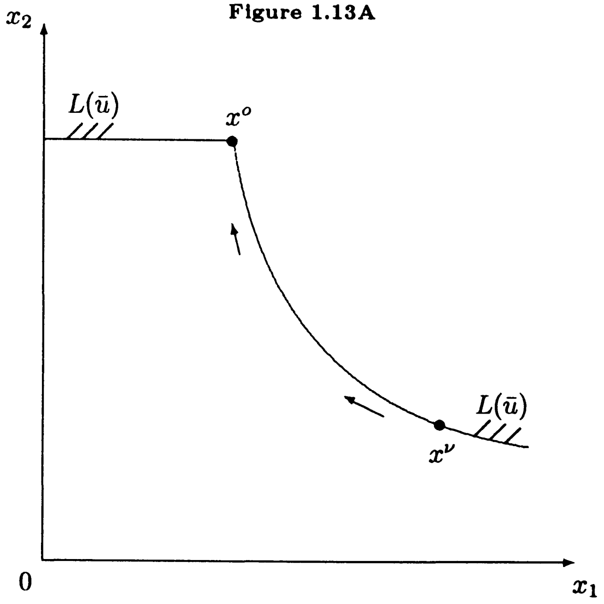

Figure 1.13B

1.3.4 Restricting the Class of Technologies

The second approach to addressing the problem posed by Bol's (1986) impossibility theorem is to restrict the class of technologies to which the efficiency index can be applied. For years, those working on the axiomatic foundations of efficiency measurement labored under the conjecture that closedness of the efficient subset of the input requirement set

$$\mathrm{Eff}(u) := \{x \in L(u) \mid \hat{x} < x \Rightarrow \hat{x} \notin L(u)\}, \tag{1.3.15}$$

was necessary and sufficient for the existence of an efficiency index that satisfied the Färe/Lovell axioms. (Note that the efficient subset in Figure 1.10 is not closed, since x^* is a limit point, but not a member, of $\mathrm{Eff}(u^*)$.) This turns out to be neither necessary nor sufficient, as shown by Dmitruk and Koshevoy (1991), who provided a complete characterization of the class of technologies for which one can construct an efficiency index satisfying the Färe/Lovell axioms (maintaining strong disposability). To state their result, define

$$A(u) = \mathrm{Eff}(u) - \mathbf{R}^n_+ \tag{1.3.16}$$

and let $\bar{A}(u)$ be the closure of $A(u)$.

(1.3.17) **Theorem**: An efficiency index satisfying (I), (M), and (H) exists for all (closed) technologies satisfying strong input disposability if and only if

$$L(u) \cap \bar{A}(u) = \mathrm{Eff}(u). \tag{1.3.18}$$

To shed some light on this condition, let us construct the requisite sets for the input requirement set, $L(u^*)$ in the Bol counterexample, reproduced in Figure 1.14. The isoquant, $I(u^*)$, is given by the entire boundary of $L(u^*)$, whereas the efficient subset, $\mathrm{Eff}(u^*)$ is given by the darkened portions of the isoquant, which excludes the flat line segment, $(\hat{x}x^*]$. The set $A(u^*)$ includes all points on or below $I(u^*)$, excluding the flat segment, $(\hat{x}x^*]$, whereas $\bar{A}(u^*)$

equals the union of A and this flat line segment. Thus, $L(u^*) \cap \bar{A}$ is coincident with $I(u^*)$ and hence is not equal to $\mathrm{Eff}(u^*)$. Thus, the Dmitruk/Koshevoy restriction rules out input requirement sets like that in Figure 1.14.

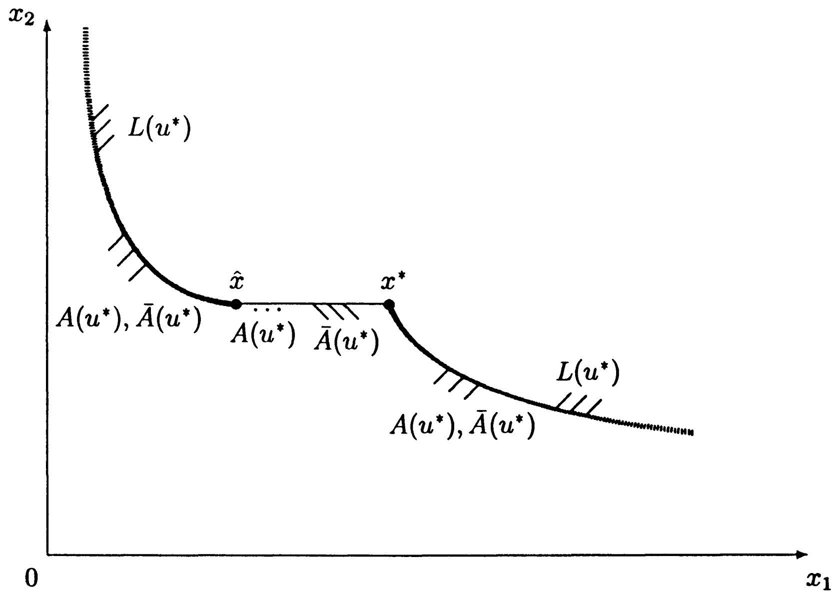

Figure 1.14

The sufficiency part of the Dmitruk/Koshevoy proof is constructive (and instructive). They construct a set $Q(u^*) \subseteq \mathbf{R}^n_+$, with the following properties:

(Q1) $Q(u^*)$ is closed.

(Q2) $Q(u^*) + \mathbf{R}^n_+ \subseteq Q(u^*)$.

(Q3) $I(Q(u^*)) = \mathrm{Eff}(u^*)$.

(Q4) $L(u^*) \subseteq Q(u^*)$.

(Q5) $\mathrm{Eff}(u^*) \subseteq \mathrm{Eff}(u^*)$.

They then show that the efficiency index $E_{DK} : \overset{\circ}{T} \longrightarrow (0,1]$, defined by

$$E_{DK}(x, u^*) := \max\{\lambda > 0 \mid x/\lambda \in Q(u^*)\}, \tag{1.3.19}$$

satisfies the Färe/Lovell conditions. Thus, the Dmitruk/Koshevoy index is simply the Debreu/Farrell index defined on the set $Q(u^*)$.

The construction of $Q(u^*)$ is a bit complicated, but essentially it is the intersection of the complements (relative to $\mathbf{R}^n_+$) of all symmetric, convex, polyhedral cones that "support" the efficient subset. To illustrate, let us return to the Färe/Lovell counterexample of Figure 1.7, reproduced in Figure 1.15 below, but with the curved portion of the isoquant replaced by a linear portion so that there exists a kink at x^*.

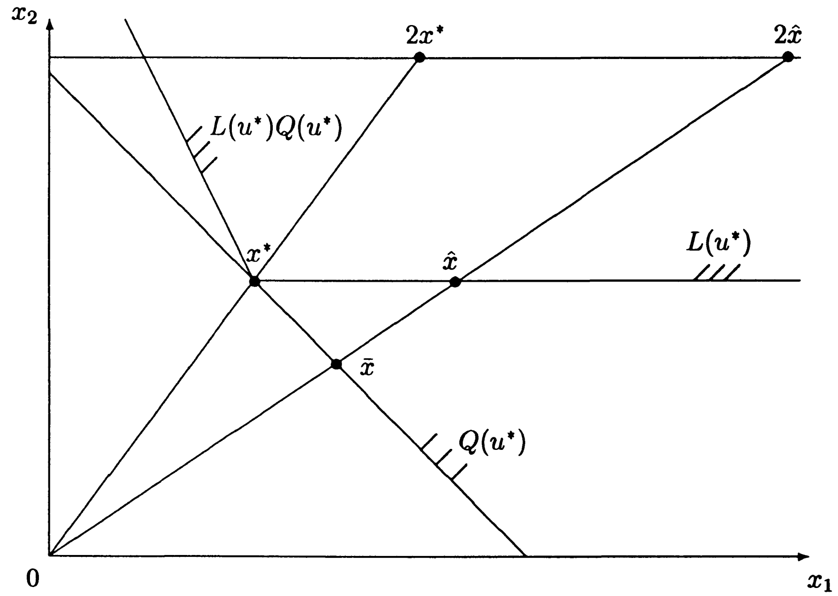

Figure 1.15

In this case, the symmetric, convex cone supporting Eff(u^*) at x^* is a half space, with the boundary given by the line, with normal $\langle 1, 1 \rangle$, through x^* and $\bar{x}$. (The supporting cones for the efficient points on the downward-sloping line segment ending at x^* are not half spaces.) The key point about the construction of $Q(u^*)$ is that, while its boundary points are coincident with the boundary points of $L(u^*)$ that are efficient (on the downward-sloping portion through x^*), the boundaries diverge where the points in $I(u^*)$ are not efficient (the flat segment through x^* and $\hat{x}$). Thus, the Debreu/Farrell efficiency index defined on $Q(u^*)$ assigns a value of $\|\bar{x}\|/\|\hat{x}\| < 1$ to $\hat{x}$, as compared to the value of 1 assigned by the Debreu/Farrell index defined on $L(u^*)$, indicating that E_{DK} satisfies the indication condition. In addition,

$$E_{DK}(u^*, 2x^*) = \|x^*\| \;/\; \|2x^*\| > E_{DK}(u^*, 2\hat{x}) = \|\bar{x}\| \;/\; \|2\hat{x}\|, \tag{1.3.20}$$

indicating that the monotonicity assumption is satisfied. Note also that, as the point $\hat{x}$ moves toward x^*, $E_{DK}(u^*, \hat{x})$ converges to 1. Finally, since the Dmitruk/Koshevoy index is just the inverse of the distance function (defined on the constructed level set $Q(u^*)$), it is homogeneous of degree minus one.

A few remarks will finish this subsection. First, compactness, but *not* closedness, of Eff(u) is sufficient for (1.3.18). Second, if T is a polyhedral set, (1.3.18) is satisfied. This is important because the technologies constructed using the mathematical programming method, or DEA, are polyhedral sets. Third, the Dmitruk/Koshevoy approach to the measurement of efficiency employs a radial, Debreu/Farrell method, obviating the need for the asymmetric coordinatewise contractions used in the Färe/Lovell and Zieschang methods.

1.3.5 Output-Based Efficiency Indexes

The treatment of the axiomatic foundations of efficiency measurement in this section has focused on the input-based efficiency index. Counterparts to all of the results exist for output-based efficiency indexes. For example, the Debreu/Farrell output-based efficiency index would simply be equal to the output distance function, D_O, restricted to the technology set. Its range is the $(0, 1]$ interval, and a value of 1 indicates output efficiency (in the weak sense that $D(u, x) = 1$ implies that u is in $\Gamma(x)$, the boundary (relative to $\mathbf{R}^m_+$) of the production possibility set $P(x)$). The Färe/Lovell concerns about the input-based Debreu/Farrell index are equally valid for the output-based counterpart, but so are other results in this section, many of which would point to the Debreu/Farrell output-based index as the appropriate concept. It must be said, however, that these results have not been formally worked out.

1.4 DUALITY AND THE DISTANCE FUNCTION

Some might complain that the material of the previous two sections contains no economics. I wouldn't argue with that characterization: indeed, much of the relevant literature appears in operations-research and management-science journals. In this section, we integrate the distance function into the core of consumer and producer theory by developing the symmetric duality between the input distance function and the cost (expenditure) function. We will also relate this duality to various index-number issues in the succeeding section.

1.4.1 The Cost (Expenditure) Function

Define

$$\bar{\mathrm{U}} = \{u \in \mathbf{R}^m_+ \mid L(u) \neq \emptyset\}. \tag{1.4.1}$$

The *cost (expenditure) function*: $C : \bar{\mathrm{U}} \times \mathbf{R}^n_{++} \to \mathbf{R}_+$, is defined by[30]

$$C(u, p) = \min_x \{p \cdot x \mid x \in L(u)\} \tag{1.4.2}$$

[30] Again, one could define the cost function for all $u \in \mathbf{R}^m_+$ (or $u \in \mathbf{R}^m$) by the convention, $C(u, p) = +\infty$ for all u such that $L(u) = \emptyset$. Alternatively, if $C(u, p)$ is defined as an infimum, the value that solves (1.4.1) when $L(u) = \emptyset$ is $+\infty$; moreover, the infimum is defined for all non-negative price vectors, $p \in \mathbf{R}^n_+$ (see Shephard (1970)).

$$
\begin{aligned}
&= \min_x \{p \cdot x \mid D_I(u,x) \geq 1\} \\
&= \min_x \{p \cdot x \mid D_O(u,x) \leq 1\} \\
&=: p \cdot \delta(u,p),
\end{aligned}
$$

where p is the price vector and δ is the vector-valued Hicksian (constant-output or constant-utility) demand function (assuming a unique solution).[31]

(1.4.3) **Theorem**: C is nondecreasing, homogeneous of degree one, concave, and continuous in p.

Thus, the cost function has monotonicity, homogeneity, curvature, and continuity properties in prices with no restrictions on the technology (other than closedness, which is needed for C to be defined).[32] The monotonicity property and homogeneity properties are obvious. Concavity is also trivial to prove, since

$$
\begin{aligned}
C(u, \alpha\hat{p} + (1-\alpha)\bar{p}) &= \min_x \left\{(\alpha\hat{p} + (1-\alpha)\bar{p}) \cdot x \mid x \in L(u)\right\} \\
&= (\alpha\hat{p} + (1-\alpha)\bar{p}) \cdot \delta(u, \alpha\hat{p} + (1-\alpha)\bar{p}) \\
&\geq \alpha C(u,\hat{p}) + (1-\alpha)C(u,\bar{p}), \qquad (1.4.4)
\end{aligned}
$$

where the last step follows directly from the definition of C; *e.g.*, $\hat{p} \cdot \delta(u, \alpha\hat{p} + (1-\alpha)\bar{p}) \geq C(u,\hat{p})$.

Stronger assumptions about the technology yield additional properties of C:

(1.4.5) **Theorem**: If the technology (or preference ordering) satisfies strong (input and output) disposability, input (preference) convexity, and continuity of L, C is (i) nondecreasing in u, (ii) nondecreasing, homogeneous of degree one, and concave in p, and (iii) jointly continuous

[31] If the solution is not unique, $\delta(u,x)$ is an arbitrary element of the demand correspondence.
[32] In fact, only closedness of $L(u)$ for all u is needed.

in u and p. Moreover, $L(u)$ is recovered from C by

$$L(u) = \bigcap_{p \gg 0^{(n)}} \{x \mid p \cdot x \geq C(u,p)\}. \tag{1.4.6}$$

The monotonicity properties follow from strong disposability. Continuity of C follows from continuity of L (most notably, lower hemicontinuity, which rules out thick isoquants or indifference curves like that in Figure 1.5B).

The recovery of $L(u)$ reflects the fact that a convex set can be characterized by the intersection of its supporting half-spaces. This concept is illustrated in Figure 1.16. Two prices, $\bar{p}$ and $\hat{p}$ generate two supporting half-spaces, $\{x \mid \bar{p} \cdot x \geq C(u^*, \bar{p})\}$ and $\{x \mid \hat{p} \cdot x \geq C(u^*, \hat{p})\}$. The intersection of these two half-spaces, with boundary defined by the darkened portions of the two lines with normals $\bar{p}$ and $\hat{p}$, provides an approximation to $L(u^*)$. Adding additional half-spaces with different normals yields closer approximations to $L(u^*)$, which is completely characterized by the intersection of such half spaces for all positive price vectors. Note that flat segments of an upper level set pose no problem for this recovery, since points like $\hat{x}$ in Figure 1.15 on page 43 would be in the intersection of the supporting half-spaces, but points below the horizontal segment of the isoquant would not. If $L(u)$ is not convex, (1.4.6) recovers the convex hull of $L(u)$; if the technology does not satisfy strong input disposability, (1.4.6) recovers the free-disposal hull of $L(u)$.[33]

Consider the special case where $m = 1$ (utility theory or single-output production) and $I(u) \cap I(\bar{u}) = \emptyset$ if $u \neq \bar{u}$. Then (1.4.2) and (1.4.6) can be replaced by

$$C(u,p) = \min_x \{p \cdot x \mid U(x) \geq u\} = p \cdot \delta(u,p) \tag{1.4.7}$$

and

$$U(x) = \max_u \{u \mid p \cdot x \geq C(u,p) \text{ for every } p \gg 0^n\}. \tag{1.4.8}$$

[33] See the references in footnotes 16 and 17 above for proofs.

This recovery of U is illustrated in Figure 1.17. Given an arbitrary x^*, assume that the "true" u^*-level set is given by $L(u^*)$. We must show that the algorithm (1.4.8) yields u^* as the maximum. Note first that u^* is feasible in the problem (1.4.8), since the constraint is satisfied as an equality, $p^* \cdot x^* = C(u^*, p^*)$, at price vector p^*, and by the definition of C, is satisfied as a weak inequality, $p \cdot x^* \geq C(u^*, p)$ for all $p \gg 0^{(n)}$. It remains to show that the constraint is violated for any $\bar{u} > u^*$. By strong output disposability, $L(\bar{u}) \subseteq L(u^*)$, as shown in Figure 1.17. Now, for some $p \gg 0^{(n)}$, such as $\bar{p}$ in Figure 1.17, the constraint in (1.4.8) is satisfied: $\bar{p} \cdot x^* > C(\bar{u}, \bar{p})$. But this constraint must be satisfied for *all* positive price vectors, and this cannot be true for p^*, the support of $L(u^*)$, since C is increasing in u $(C(\bar{u}, p^*) > C(u^*, p^*) = p^* \cdot x^*)$.

The foregoing algorithms for recovering the technology or preference ordering from the cost function, while standard components of duality theory for many years, are not operational, since they entail an infinite number of constraints. An ingenious construction by Primont and Sawyer (1993), however, provides an operational recovery mechanism. In the utility or single-output case, we have the following result:

(1.4.9) **Theorem**: Suppose $m = 1$ and the technology (or preference ordering) satisfies strong (input and output) disposability, and input (preference) convexity. Then U is recovered by

$$U(x) = \inf_{u,p} \{u \mid p \cdot x \leq C(u, p)\}. \tag{1.4.10}$$

The beauty of this result is that it is a fairly standard optimization problem with a single constraint. If C is differentiable, (1.4.10) can be used to recover a utility or production function by straightforward calculation methods. Even without differentiability, it is a useful mechanism for proving properties of utility or production functions from given properties of cost functions.[34]

[34] Blackorby and Russell (1997) have recently made good use of this mechanism.

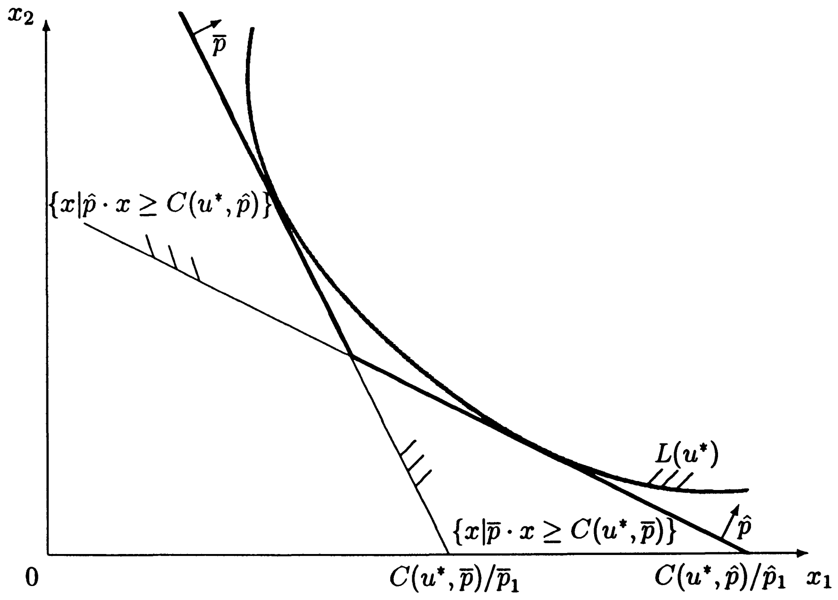

Figure 1.16

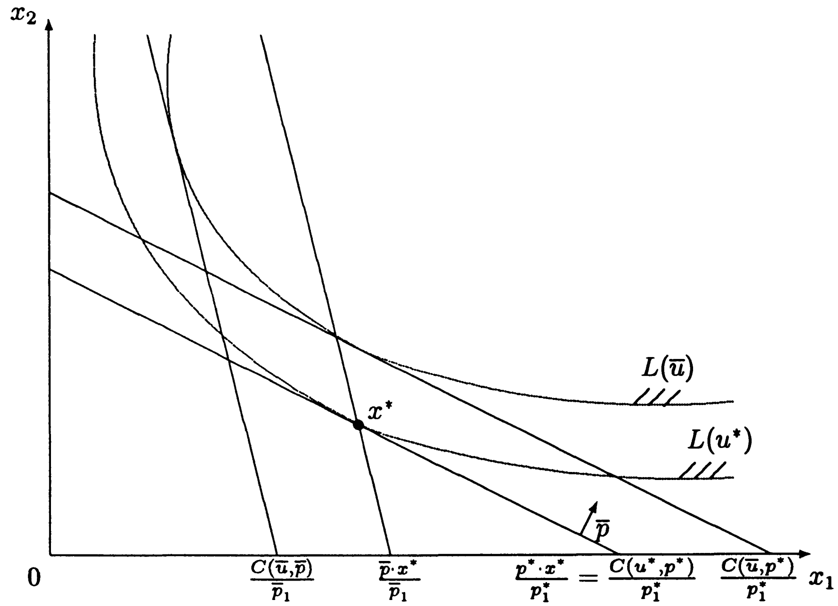

Figure 1.17

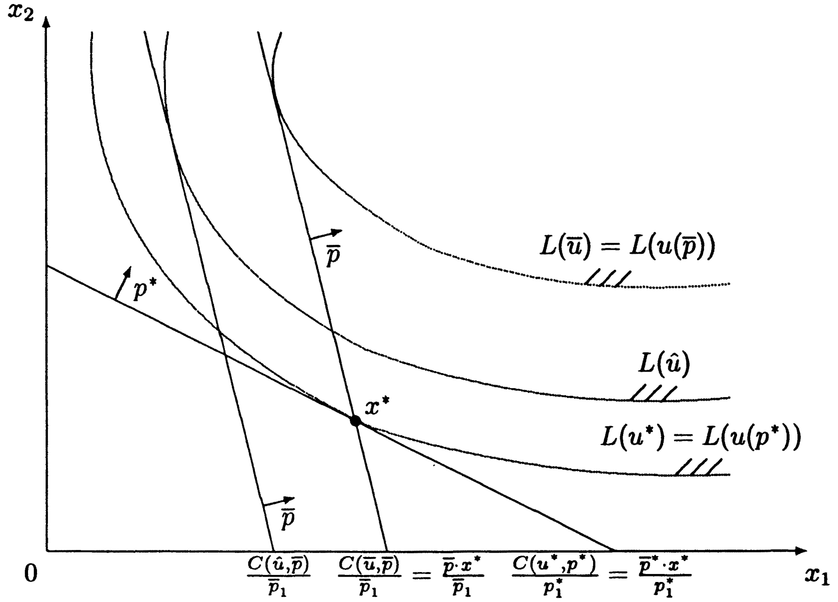

Figure 1.18

The intuition for the Primont/Sawyer result can be illustrated in two dimensions. Suppose that we are given a cost function C and a quantity vector x^* and suppose that $L(u^*)$ in Figure 1.18 represents the truth about U. Then, for a given $\bar{p}$, $\bar{u}$ solves the problem,

$$\min\{u \mid \bar{p} \cdot x^* \leq C(u, \bar{p})\} =: u(\bar{p}). \tag{1.4.11}$$

The constraint in (1.4.10) is satisfied as an equality, $\bar{p} \cdot x^* = C(\bar{u}, \bar{p})$ and, for any smaller u, say $\hat{u}$, we have $L(\bar{u}) \subset L(\hat{u})$, with nonintersecting isoquants, so that $\bar{p} \cdot x^* > C(u, \bar{p})$. Note that $\theta\bar{p}$ also solves (1.4.11), since $\theta\bar{p} \cdot x^* = C(\bar{u}, \theta\bar{p})$, because of the first-degree homogeneity of C in p. Thus, given a solution, $u(p)$ to the utility minimization problem (1.4.11) for any p, (1.4.10) is solved by choosing the utility-minimizing direction of p:

$$\min_{p(u)}\{u \mid p(u) \cdot x^* \leq C(u, p(u))\}. \tag{1.4.12}$$

The solution is clearly given by p^*, the support of $L(u^*)$ at x^*, since any other price direction, such as $\bar{p}$, would yield a higher minimal utility subject to the constraint in (1.4.12).

Primont and Sawyer extended their recovery mechanism to multiple-output technologies (under somewhat different regularity conditions) by recovering the *output* distance function:[35]

(1.4.13) **Theorem**: If the technology satisfies strong output disposability and input convexity and C is derived from D_O by (1.4.2), then for $\langle u, x \rangle$ such that x is cost minimizing for some p, D_O is recovered by

$$D_O(u, x) = \sup_{p,\lambda} \{\lambda \mid C(u/\lambda, p) \geq p \cdot x\}. \qquad (1.4.14)$$

While the intuition for this multiple-output result is a little more complicated, a graphical illustration is possible. In Figure 1.19, two output quantities are measured in the third quadrant, while two input quantities are measured in the first quadrant. Without loss of generality, consider $\langle \bar{u}, \bar{x} \rangle$ such that $\bar{u} \in P(\bar{x})$ but $\bar{u} \notin \Gamma(\bar{x})$. Then $D_O(\bar{u}, \bar{x}) = \|\bar{u}\| \; / \; \|\hat{u}\| = \lambda^* < 1$. By strong input disposability, $L(\bar{u}/\lambda^*) \subseteq L(\bar{u})$ and $\bar{x} \in I(\hat{u}) = I(\bar{u}/\lambda^*)$. Now fix p at $\hat{p}$ and find

$$\lambda(\hat{p}) = \max\{\lambda \mid C(\bar{u}/\lambda, \hat{p}) \geq \hat{p} \cdot \bar{x}\}. \qquad (1.4.15)$$

This maximum is attained where the constraint is satisfied as an equality, $C(\bar{u}/\lambda(\hat{p}), \hat{p}) = \hat{p} \cdot \bar{x}$, as shown in Figure 1.19. Any larger value of λ would result in an inequality that violated the constraint in (1.4.15). Now choose the price (direction) to solve

$$\max_{p \gg 0^{(n)}} \{\lambda(p) \mid C(\bar{u}/\lambda(p), p) \geq p \cdot \bar{x}\}. \qquad (1.4.16)$$

The maximum is reached at $\bar{p}$. Any other price direction, such as $\hat{p}$, results in a higher $u/\lambda(p)$, and hence a lower value of $\lambda(p)$.

[35] Note that the Primont/Sawyer assumption in this theorem that C is increasing along rays in u (*i.e.*, $C(tu, p) > C(u, p) \; \forall \; t > 1$) is implied by our strong output disposability assumption. Cf. equation (1.4.29) below.

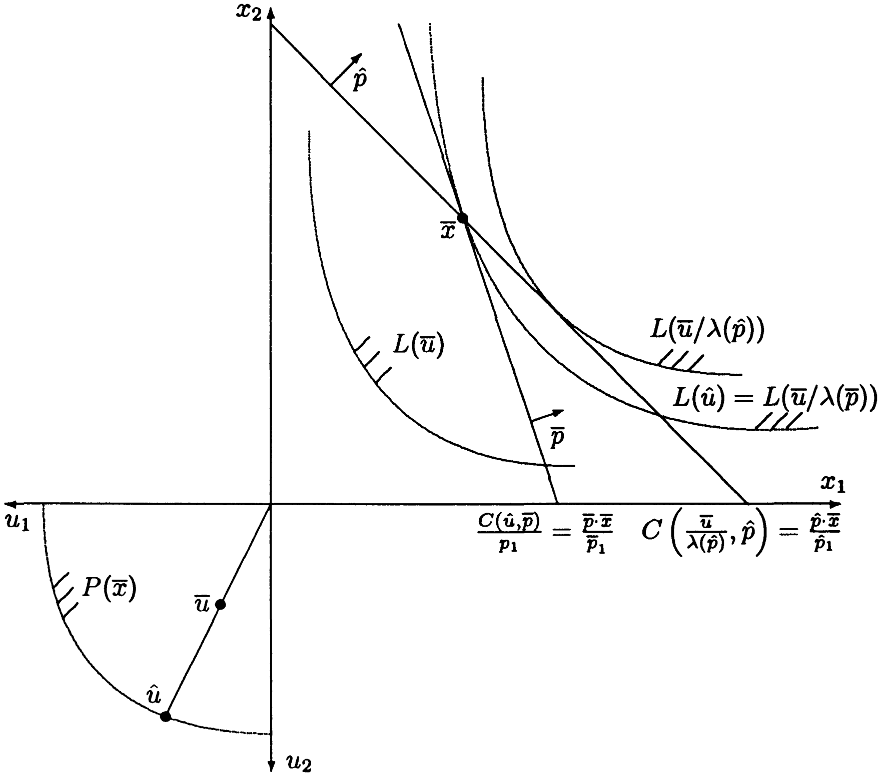

Figure 1.19

Starting with a technology or preference ordering, deriving a cost (expenditure) function, and then recovering the technology or preference ordering is an important aspect of duality theory; closely related but perhaps more important for empirical research is the problem characterized by the following:

(1.4.17) **Corollary**: Given a C that satisfies (i), (ii), and (iii) in Theorem 1.4.5, there exists a technology (or preference ordering), T, satisfying strong (input and output) disposability, convexity of $L(u)$ for all u, and continuity of L, such that C is derived from T by (1.4.2).

Thus, specifying a cost function C that satisfies the monotonicity, homogeneity, curvature, and continuity properties in Theorem 1.4.5 is tantamount to specifying a technology or preference ordering with the requisite properties of the theorem such that the specified cost function is generated by (1.4.2).

1.4.2 Symmetric Duality Between the Distance and Cost (Expenditure) Functions

The symmetry between the input distance function and the cost function is instructive: given certain regularity conditions for the technology or preference ordering, D_I and C have identical properties—positive monotonicity, concavity, and first-degree homogeneity—in x and p, respectively. Moreover, each is monotonic in u (though in opposite directions) and each is jointly continuous in $\langle u, x\rangle$. The symmetry does not end there. It turns out, unsurprisingly, that the cost function is a distance function in price space and the input distance function is an imputed cost function in price space. Moreover, each is derived from the other by identical optimization problems.

To illustrate the first property, consider the u-dependent price sets,

$$L^*(u) = \{p \in \mathbf{R}^n_{++} \mid C(u,p) \geq 1\}. \tag{1.4.18}$$

These upper-level sets for C can be interpreted as containing the price vectors that yield a minimal cost of producing output vector u, or utility level u, that is no less than 1. As such, their interiors are the sets of prices that are too high to produce output u or attain utility level u at expenditure level 1. Computing these level sets at unit expenditure is simply a convenient and harmless normalization. One could alternatively define level sets

$$L^*(u) = \{p/y \in \mathbf{R}^n_{++} \mid C(u,p) \geq y\} = \{p/y \mid C(u,p/y) \geq 1\}, \tag{1.4.19}$$

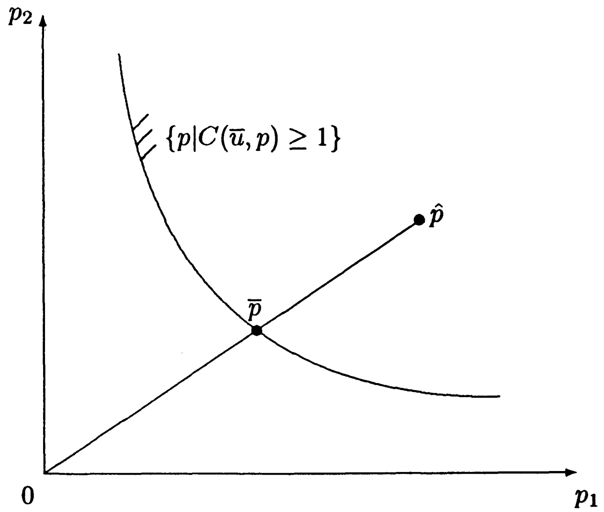

Figure 1.20

which would identify the set of normalized prices that are too high, or just low enough, to obtain output or utility u. One such level set is drawn in Figure 1.20. In the case of utility theory, these sets are equal to $\{p \in \mathbf{R}^n_{++} \mid V(p) \geq \bar{u}\}$, where V is the indirect utility function, and the boundaries of such level sets are indirect indifference curves (with higher indifference curves corresponding to lower utility levels).

Now define the indirect distance function, $D_I^* : \bar{\mathbf{U}} \times \mathbf{R}^n_{++} \to (0, +\infty)$:

$$D_I^*(u,p) = \max\{\lambda > 0 \mid C(u, p/\lambda) \geq 1\}. \tag{1.4.20}$$

In Figure 1.20, $D_I^*(\bar{u},\hat{p}) = \|\hat{p}\| / \|\bar{p}\|$. As C is homogeneous of degree one in p, it is trivial to show that

$$\begin{aligned} D_I^*(u,p) &= \max\{\lambda > 0 \mid C(u,p/\lambda) \geq 1\} \\ &= \max\{\lambda > 0 \mid C(u,p) \geq \lambda\} \\ &= C(u,p). \end{aligned} \tag{1.4.21}$$

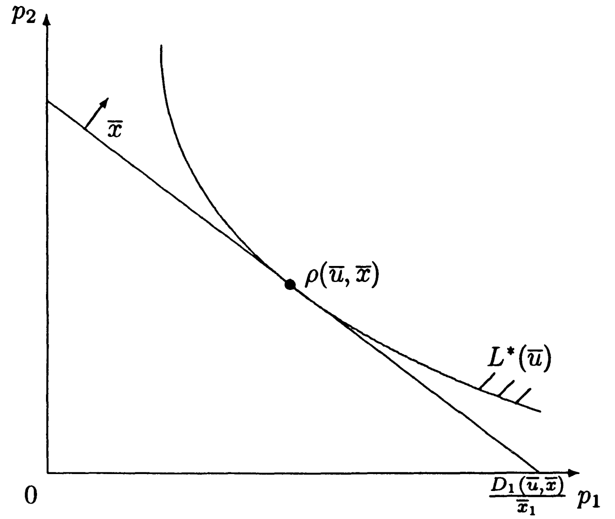

Figure 1.21

Thus, the cost function is a distance function in price space.

We now demonstrate that, symmetrically, the input distance function is a cost imputation function in price space (assuming input convexity and strong input disposability); that is,

$$D_I(u,x) = \inf_{p \gg 0^{(n)}} \{p \cdot x \mid C(u,p) \geq 1\} = \rho(u,x) \cdot x, \qquad (1.4.22)$$

where $\rho(u,x)$ is the argmin of this optimization problem (assuming it's unique). This construction is illustrated in Figure 1.21. Thus, $\rho(\bar{u},\bar{x})$ is the vector of (shadow) prices that minimizes the imputed value of quantity vector $\bar{x}$ subject to the constraint that prices be high enough that the minimum cost of producing $\bar{u}$ is no less than 1. (Note that $C(u,\rho(u,x)) = 1$.) Alternatively, $\rho(u,p)$ can be interpreted as the vector of shadow prices deflated by total expenditure, since homogeneity of C implies that $C(u,p) = y$ if and only if $C(u,p/y) = 1$; that is, the cost minimization problem is the same at income 1 and price vector p/y as at income y and price

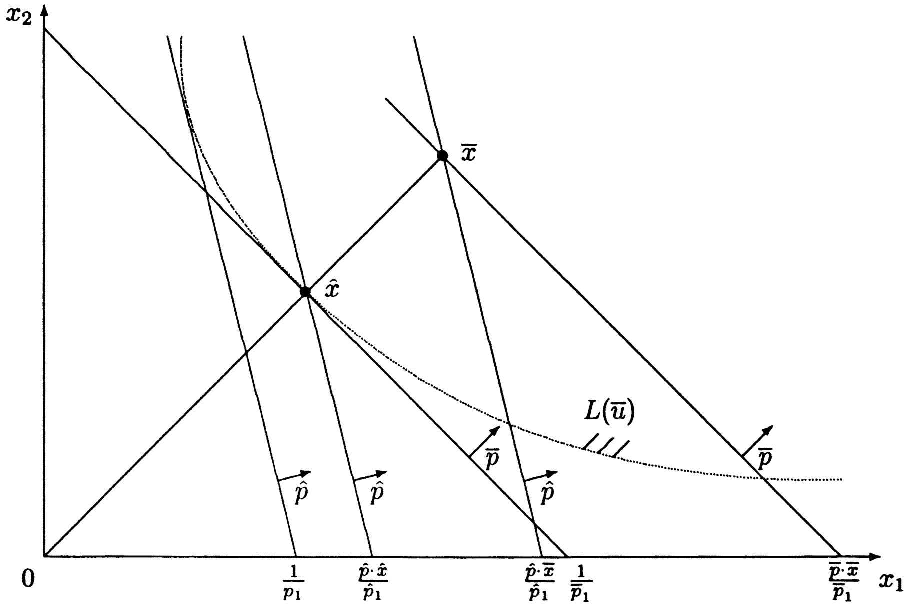

Figure 1.22

vector p.

To demonstrate this characterization of the input distance function, let us return to quantity space in Figure 1.22. We are given an output quantity vector or utility scalar $\bar{u}$ and an input or consumption vector $\bar{x}$. The "truth" is known: $D_I(\bar{u}, \bar{x}) = ||\bar{x}|| \ / \ ||\hat{x}||$; we must show that the optimization problem in (1.4.22) yields the same answer. In particular, we will show that $\bar{p}$, the support of $L(\bar{u})$ at $\hat{x}$, is the only possible solution to (1.4.22). If $\bar{p}$ solves (1.4.22), then $C(\bar{u}, \bar{p}) = 1$ and

$$\frac{\bar{p}\cdot\bar{x}}{C(\bar{u},\bar{p})} = \frac{\bar{p}\cdot\bar{x}}{1} = \frac{||\bar{x}||}{||\hat{x}||} = D_I(\bar{u},\bar{x}). \qquad (1.4.23)$$

Now, any other candidate for a solution cannot have the same direction as $\bar{p}$, since the solution satisfies $C(u, p) = 1$. Consider,

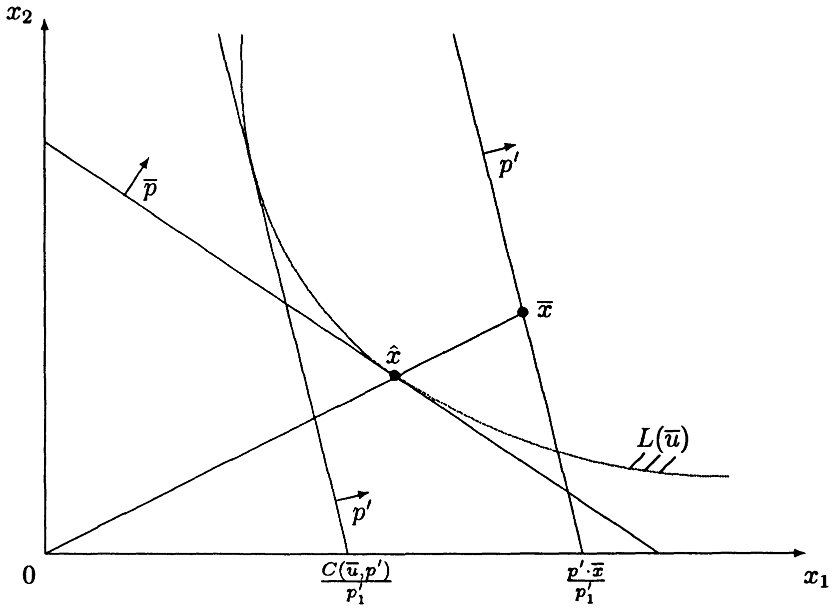

Figure 1.23

therefore, $\hat{p}$ as a solution and note that

$$\frac{\hat{p}\cdot\bar{x}}{\hat{p}\cdot\hat{x}} = \frac{||\bar{x}||}{||\hat{x}||} = \bar{p}\cdot\bar{x}. \tag{1.4.24}$$

But since $\hat{p}\cdot\hat{x} > C(\bar{u},\hat{p}) = 1$, it follows that $\hat{p}\cdot\bar{x} > \bar{p}\cdot\bar{x}$, indicating that the alternative price vector yields a value of the objective function in (1.4.22) that is higher than $\bar{p}\cdot\bar{x}$. Thus, the input distance function is a cost imputation function. Moreover, as $\bar{p} = \rho(\bar{u},\bar{x})$ is the support of $L(\bar{u})$ in the direction $\bar{x}$, it is clear that $\rho(u,x)$ is a vector of shadow prices but only its direction is meaningful, since it is only the ratio of $\bar{p}_1/\bar{p}_2$ that determines the solution in Figure 1.22. That is, if $\bar{p}$ is a solution, so is $\tilde{p} = \theta\bar{p}$ with expenditure equal to $\theta > 0$.

The duality between the input distance function and cost function is further explicated in Figure 1.23, in which

$$D_I(\bar{u},\bar{x}) = \frac{\|\bar{x}\|}{\|\hat{x}\|} < \frac{p' \cdot \bar{x}}{C(\bar{u},p')}. \tag{1.4.25}$$

Moreover, this inequality will hold for any price other than $\bar{p}$, which supports $L(\bar{u})$ at $\hat{x} = \delta(\bar{u},\bar{p})$; in this case

$$D_I(\bar{u},\bar{x}) = \frac{\bar{p} \cdot \bar{x}}{C(\bar{u},\bar{p})}. \tag{1.4.26}$$

Price and quantity vectors satisfying this equality are called "conjugate pairs" (see Gorman (1976)). Thus, we have "Mahler's inequality":

$$D_I(u,x)C(u,p) \leq p \cdot x \tag{1.4.27}$$

for all triples $\langle u,x,p\rangle$ in the domains of D_I and C, with equality holding if x and p are a conjugate pair relative to u. It is also clear from Mahler's inequality, and from Figure 1.23, that the constrained optimization problems (1.4.2) and (1.4.22) can be re-written as unconstrained optimization problems,

$$C(u,p) = \min_x \left(\frac{p \cdot x}{D_I(u,x)}\right) \tag{1.4.28}$$

and

$$D_I(u,x) = \inf_p \left(\frac{p \cdot x}{C(u,p)}\right). \tag{1.4.29}$$

(Note that, at the optima, $D_I(u,\delta(u,p)) = 1$ and $C(u,\rho(u,x)) = 1$.)

The symmetry of the duality between the input distance function and the cost (expenditure) function is underscored by the following summary of the derivation of each from the other:

$$C(u,p) = \min_x \left\{p \cdot x \mid D_I(u,x) \geq 1\right\} = \min_x \left(\frac{p \cdot x}{D_I(u,x)}\right) \tag{1.4.30}$$

and

$$D_I(u,x) = \inf_p \left\{p \cdot x \mid C(u,p) \geq 1\right\} = \inf_p \left(\frac{p \cdot x}{C(u,p)}\right), \tag{1.4.31}$$

which makes palpable the symmetry between the two representations of technologies or preferences (under the appropriate regularity conditions).

1.4.3 Shephard's Lemma and its Dual

It is now textbook knowledge that much of the power in the theoretical and empirical application of the duality between production technologies or preference orderings and the cost or expenditure function flows through the use of

Shephard's Lemma: If C is differentiable in p at $\langle u, p\rangle$,

$$\delta(u,p) = \nabla_p C(u,p); \tag{1.4.32}$$

i.e.,

$$\delta_i(u,p) = \frac{\partial C(u,p)}{\partial p_i}, \; i = 1, \ldots, n, \tag{1.4.33}$$

which says that Hicksian (constant-output or constant-utility) demand functions are generated straight away by simple differentiation of the cost function. As is now well known, this celebrated result is a straightforward application of the envelope theorem to the optimization problem (1.4.2).

Many theoretical and applied issues are best analyzed—or often can only be analyzed—through the use of constant-output or income-compensated (Hicksian) demand functions. Indeed, the entire (differential) comparative-static content of neoclassical consumer theory follows trivially from duality theory using Shephard's Lemma (assuming twice differentiability of C), which implies that the $\langle i, j\rangle$ (Slutsky) substitution effect is simply

$$S_{ij}(u,p) := \frac{\partial \delta_i(u,p)}{\partial p_j} = C_{ij}(u,p), \tag{1.4.34}$$

where C_{ij} is the cross partial derivative of the expenditure function with respect to prices. Thus, the symmetry, negative semi-definiteness, and reduced rank of the Slutsky substitution matrix,

$$S(u,p) := \nabla_{pp}C(u,p) = \begin{bmatrix} C_{11}(u,p) & C_{12}(u,p) & \dots & C_{1n}(u,p) \\ C_{21}(u,p) & C_{22}(u,p) & \dots & C_{2n}(u,p) \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1}(u,p) & C_{n2}(u,p) & \dots & C_{nn}(u,p) \end{bmatrix}, \tag{1.4.35}$$

follows immediately from Young's Theorem, concavity of C, and homogeneity of degree one of C (homogeneity of degree zero of C_i for all i).[36] As shown on page 46, proofs of concavity and homogeneity of C are trivial. Moreover, these properties of the Slutsky substitution matrix hold whether or not preferences are convex (see Theorem 1.4.3).[37] Conversely, an arbitrary system of Hicksian demand functions can be integrated to an (indirect) utility function if and only if $S(u,p)$ satisfies these properties (see Samuelson (1947, 1950)).

In addition to the trivial derivation of the comparative-static content of consumer theory, the examination of many other issues involving the properties of isoquant surfaces in production theory (*e.g.*, elasticities of substitution[38]) or of income-compensated demands (*e.g.*, consumer surplus[39]) are facilitated by the use of Shephard's Lemma.

The application of the envelope theorem to the optimization problem (1.4.22), yields the

[36]Homogeneity of degree zero of C_i in p implies that $\sum_j C_{ij}(u,p)p_j = 0$, so that the columns of $S(u,p)$ are linearly dependent.

[37]The Slutsky decomposition of the price gradient of the ordinary (Marshallian) demand functions into an income effect and a substitution effect is also trivial using Shephard's Lemma. (See, *e.g.*, Varian (1992).)

[38]See Blackorby and Russell (1976, 1981)

[39]See, *e.g.*, Varian (1992).

Dual to Shephard's Lemma: If D_I is differentiable in x at $\langle u, x \rangle$,

$$\rho(u,x) = \nabla_x D_I(u,x); \tag{1.4.36}$$

i.e.,

$$\rho_i(u,x) = \frac{\partial D_I(u,x)}{\partial x_i}, \; i = 1, \ldots, n, \tag{1.4.37}$$

which generates shadow price functions, $\rho_i(u,x)$, $i = 1, \ldots, n$, as simple derivatives of the (input) distance function. As D_I is homogeneous of degree one in x, ρ is homogeneous of degree zero in x; hence, the value of this shadow-price vector depends only on the direction of the quantity vector x. This fact is readily apparent from the illustration of the optimization problem (1.4.22) in Figure 1.24B, where x' is the normal of the hyperplane that supports $L^*(u')$ at $\rho(u', x')$. The dual representation, in Figure 1.24A, indicates that $\rho(u', x')$ is the support of $L(u')$ (the marginal rate of substitution) at $x'/D_I(u', x')$; clearly, this is also the support at $\lambda x'/D_I(u', \lambda x')$ for any $\lambda > 0$. Recall from Subsection 1.4.2 that $\rho(u,x)$ is the vector of shadow prices normalized at unit expenditure or, equivalently, the vector of absolute shadow prices divided by total expenditure. If x is not cost-minimizing at prices $\rho(u,x)$ and utility u, $\rho(u,x)$ must be interpreted as the shadow price vector deflated by $C(u, \rho(u,x))$.

The comparative statics of shadow prices, or inverse Hicksian demands, is embodied in the properties of

$$\frac{\partial \rho_i(u,x)}{\partial x_j} =: D_{ij}(u,x) \tag{1.4.38}$$

where D_{ij} is the second-order cross derivative of D_I with respect to quantities. The comparative statics are thus characterized by the properties of the Hessian of the input distance function,

$$A(u,x) := \nabla_{xx} D_I(u,x) = \begin{bmatrix} D_{11}(u,x) & D_{12}(u,x) & \ldots & D_{1n}(u,x) \\ D_{21}(u,x) & D_{22}(u,x) & \ldots & D_{2n}(u,x) \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}(u,x) & D_{n2}(u,x) & \ldots & D_{nn}(u,x) \end{bmatrix} \tag{1.4.39}$$

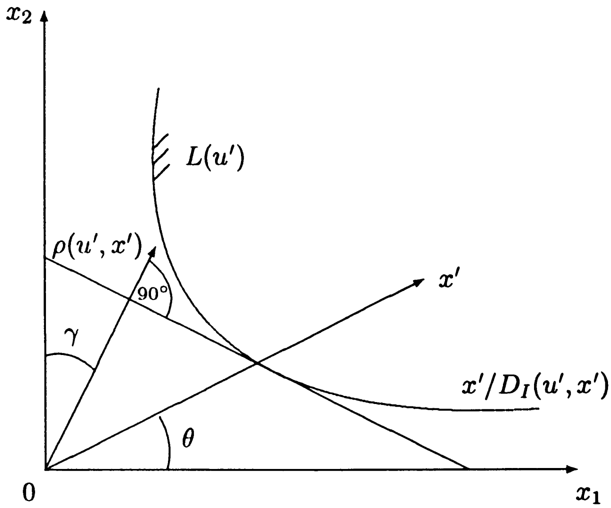

Figure 1.24A

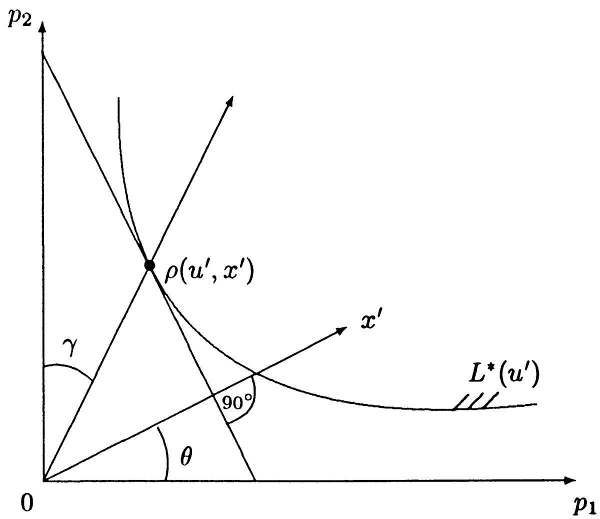

Figure 1.24B

Young's Theorem and homogeneity of D_I guarantee that $A(u,x)$ is symmetric and has reduced rank. If technologies or preferences are convex and satisfy strong disposability (or strong nonsatiation), then by Theorem 1.2.28 $A(u,x)$ is also negative semi-definite.

Thus, under the appropriate convexity and disposability (nonsatiation) assumptions, $A(u,x)$ has the same properties as the Slutsky matrix. The latter, however, has these properties whether or not preferences or technologies are convex and satisfy strong nonsatiation or strong disposability. The reason for this distinction is that the (primitive) definition of D_I in (1.2.18) takes no account of optimizing behavior. If D_I is *defined* as a cost-imputation function, as in (1.4.22), then $A(u,x)$ is negative semi-definite whether or not preferences or technologies are nonconvex or violate strong nonsatiation. In analyzing behavior, as opposed to providing a non-behavioral characterization of a preference ordering or technology, it makes sense to define the distance function by (1.4.22).[40] In this case, the comparative statics of shadow-price determination are completely symmetric to the comparative statics of Hicksian demand, and the symmetry, negative semi-definiteness, and reduced rank of $A(u,x)$ completely summarizes the comparative statics of consumer theory (and the theory of constant-output demand of producers). Moreover, these properties of $A(u,x)$ are necessary and sufficient for the integrability of an inverse Hicksian demand system to a "behavioral distance function." Indeed, the integrability problem was originally framed by Antonelli in terms of inverse demand functions; thus, Samuelson (1947, 1950) refers to $A(u,x)$ as the "Antonelli matrix."[41]

[40]Note that data on observed behavior could never detect violations of the convexity or strong disposability (nonsatiation) assumptions.

[41]As shown by Deaton (1978), the Slutsky and Antonelli matrices are generalized inverses of one another.

On the other hand, in some situations it might be inappropriate to assume cost-minimizing, price-taking behavior. This could occur because of allocative inefficiency within firms or because of market inefficiencies attributable to imperfect competition or regulation in input markets.[42] In this case, the appropriate concept for calculating shadow prices is the primitive notion of (1.4.22).

The derivation of shadow-price functions from the (input) distance function has not yet been extensively exploited. The principal applications of which I am aware are to optimal tax theory (Deaton (1978, 1981)) and to the determination of shadow prices in regulated and other non-competitive environments (Färe, Grosskopf, and Nelson (1990) and Färe and Grosskopf (1990)).

1.5 DISTANCE FUNCTIONS AND INDEX NUMBERS

1.5.1 The Konüs Cost-of-Living Index

The Könus (1924) "true" cost-of-living index, which is based on utility theory (unlike "mechanistic" price indexes[43]), is defined as the ratio of the (minimal) cost of obtaining utility level u at situation t prices to the cost of obtaining the same utility at situation t' prices. We might think of situation t as the "current" period and situation t' as the "base" period, but the situations could refer to different locations or other types of scenarios as well.

Formally, define the *Konüs cost-of-living index* by

$$COL^K(u,p^t,p^{t'}) = \frac{C(u,p^t)}{C(u,p^{t'})}. \tag{1.5.1}$$

[42]Overall cost inefficiency of (price-taking) firms, $C(u,p)/p \cdot x$, can be decomposed into (Debreu/Farrell) technical inefficiency, as defined by (1.3.1), and allocative inefficiency, $p \cdot (x/D_I(u,x))/C(u,p)$. See Farrell (1957) and Färe, Grosskopf, and Lovell (1994).

[43]See Eichhorn and Voeller (1976) and Balk (1995) for excellent treatments of the mechanistic (axiomatic) approach to index numbers.

In this formulation, u is the "reference" utility level. In many cases, it is natural to set u equal to the current-period or base-period utility, in which case

$$COL^K(u^t, p^t, p^{t'}) = \frac{C(u^t, p^t)}{C(u^t, p^{t'})} \tag{1.5.2}$$

or

$$COL^K(u^{t'}, p^t, p^{t'}) = \frac{C(u^{t'}, p^t)}{C(u^{t'}, p^{t'})}, \tag{1.5.3}$$

where $u^t = V(p^t/y^t)$ and y^t is total expenditure in situation t.

Thus,

$$\left(COL^K(u^{t'}, p^t, p^{t'}) - 1\right) \times 100$$

tells us the percentage increase in total expenditure between period t' and period t needed to allow the consumer to obtain the same utility in period t that he or she enjoyed in period t'.

Using Shephard's decomposition theorem,[44]

[44]Proof can be found in most treatments of duality theory, but it is easy. Since homotheticity of U is equivalent to $U(x) = \Phi\big(\bar{U}(x)\big)$, where Φ is increasing and $\bar{U}$ is homogeneous of degree one,

$$\begin{aligned} C(u,p) &= \min_x \left\{p \cdot x \;\middle|\; \Phi\left(\bar{U}(x)\right) \geq u\right\} \\ &= \min_x \left\{p \cdot x \;\middle|\; \bar{U}(x) \geq \Phi^{-1}(u)\right\} \\ &= \Phi^{-1}(u) \min_{x/\Phi^{-1}(u)} \left\{p \cdot \frac{x}{\Phi^{-1}(u)} \;\middle|\; \bar{U}\left(\frac{x}{\Phi^{-1}(u)}\right) \geq 1\right\} \\ &= \Phi^{-1}(u)\; C(1,p) =: \Psi(u)\; \Pi(p). \end{aligned}$$

Conversely,

$$\begin{aligned} DI(u,x) &= \inf_p \{p \cdot x \,|\, \Phi(u)\Pi(p) \geq 1\} \\ &= \frac{1}{\Psi(u)} \inf_{\Psi(u)\cdot p} \{\Psi(u)p \cdot x \,|\, \Pi(\Psi(u)p) \geq 1\} \\ &= \frac{D(1,x)}{\Psi(u)} =: \frac{\Lambda(x)}{\Psi(u)}, \end{aligned}$$

where Λ is homogeneous of degree one. Inverting $\Lambda(x)/\Psi(u) = 1$ in u yields a homothetic utility function.

$$U \text{ is homothetic} \iff C(u,p) = \Psi(u)\ \Pi(p), \tag{1.5.4}$$

where Ψ is increasing and Π is nondecreasing, concave, and homogeneous of degree one, we have the following well-known result:

(1.5.6) **Theorem**: $COL^K(u, p^t, p^{t'})$ is independent of u if and only if preferences are homothetic, in which case

$$COL^K(u, p^t, p^{t'}) = \frac{\Pi(p^t)}{\Pi(p^{t'})}. \tag{1.5.7}$$

Especially in the mechanistic price theory literature, cost-of-living indexes are often ratios of weighted averages of prices in the two situations:

$$COL^K(u, p^t, p^{t'}) = \frac{\sum_i a_i p_i^t}{\sum_i a_i p_i^{t'}}. \tag{1.5.8}$$

This index can be rationalized as a Konüs cost-of-living index if preferences are Leontief:

$$U(x) = \min\left\{\frac{x_1}{a_1}, \ldots, \frac{x_n}{a_n}\right\}. \tag{1.5.9}$$

In this case

$$\begin{aligned} C(u,p) &= \min_x \left\{ p \cdot x \;\middle|\; \min\left\{\frac{x_1}{a_1}, \cdots, \frac{x_n}{a_n}\right\} \geq u \right\}, \qquad (1.5.10) \\ &= u \sum_i a_i\, p_i, \end{aligned}$$

and taking the appropriate ratio yields (1.5.8).

It is sometimes said that the Leontief preferences given by (1.5.9) are *necessary* as well as sufficient for a Konüs cost-of-living index to be a ratio of linear functions of prices in the two situations. This is not so, and the confusion derives from contemplating only explicit utility functions to accommodate the lack of a substitution effect implicit in these indexes. A complete characterization using the distance function is provided by the following theorem:

(1.5.11) **Theorem**: The Konüs cost-of-living index is a ratio of weighted averages in prices if and only if preferences are represented by the "implicit Leontief function":

$$D_I(u,x) = \min\left\{\frac{x_1}{a_1(u)}, \cdots, \frac{x_n}{a_n(u)}\right\}, \tag{1.5.12}$$

where the a_i, $i = 1, \ldots, n$, are positive nondecreasing functions, in which case,

$$\begin{aligned} C(u,p) &= \min_x \left\{p \cdot x \;\middle|\; \min\left\{\frac{x_1}{a_1(u)}, \cdots, \frac{x_n}{a_n(u)}\right\} \geq 1\right\} \\ &= \sum_i a_i(u)\, p_i, \end{aligned} \tag{1.5.13}$$

so that

$$COL^K(u, p^t, p^{t'}) = \frac{\sum_i a_i(u)\, p_i^t}{\sum_i a_i(u)\, p_i^{t'}}. \tag{1.5.14}$$

A two-dimensional example of the implicit Leontief preference ordering is given is Figure 1.25. Note that (1.5.12) cannot be solved for a closed form utility function unless each function, a_i is linear, in which case (1.5.13) simplifies to (1.5.10) and (1.5.14) simplifies to (1.5.9).

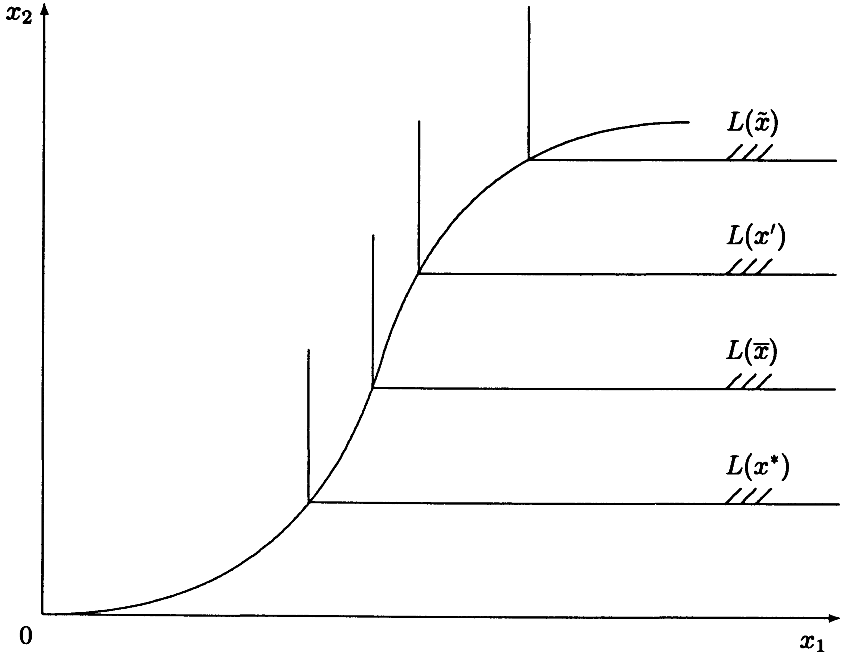

Figure 1.25

Special cases of the linear cost-of-living indexes with a long and venerable history are the Paasche and Laspeyres price indexes, defined by

$$COL^P(x^t, p^t, p^{t'}) = \frac{\sum_i x_i^t \, p_i^t}{\sum_i x_i^t \, p_i^{t'}} \tag{1.5.15}$$

and

$$COL^L(x^{t'}, p^t, p^{t'}) = \frac{\sum_i x_i^{t'} \, p_i^t}{\sum_i x_i^{t'} \, p_i^{t'}}, \tag{1.5.16}$$

where t and t' are definitively the current period and base period, respectively. The following result rationalizes these indexes in terms of utility theory:

(1.5.18) **Theorem**: If the consumer is cost minimizing and utility is given by the implicit Leontief (1.5.12), then

$$COL^K(u^t, p^t, p^{t'}) = COL^P(x^t, p^t, p^{t'}) \tag{1.5.19}$$

and

$$COL^K(u^{t'}, p^t, p^{t'}) = COL^L(x^{t'}, p^t, p^{t'}). \qquad (1.5.20)$$

This result is easy to see since, at the cost-minimizing bundles, x^t and $x^{t'}$, $D(u^t, x^t) = 1$ and $D(u^{t'}, x^{t'}) = 1$, so that

$$\frac{x_i^t}{a_i(u^t)} = 1 \quad \forall\, i \quad \text{and} \quad \frac{x_i^{t'}}{a_i(u^{t'})} = 1 \quad \forall\, i. \qquad (1.5.21)$$

Substitution of these identities, along with u^t and $u^{t'}$, respectively, into (1.5.14) yields (1.5.15) and (1.5.16).

The interesting aspect of Theorem 1.5.18 is that the Paasche and Laspeyres indexes are completely data based in that they depend only on observable quantities and prices in the two situations and require no estimation of preference parameters, yet are rationalized by a well-behaved utility function—that is, they are "exact" for the implicit Leontief preferences given by (1.5.12). On the other hand, these utility rationalizations of the Paasche and Laspeyres indexes are not especially exciting, since (1.5.12) does not represent a very robust class of preference orderings. As is well known, they allow for no net substitutability between commodities, although there is gross substitutability attributable to the income effect. The following bounds do offer some justification for the Paasche and Laspeyres indexes, at least when the bounds are fairly tight:[45]

$$COL^P(x^t, p^t, p^{t'}) \leq COL^K(u^t, p^t, p^{t'}) \leq \max_i \left\{ \frac{p_i^t}{p_i^{t'}} \right\} \qquad (1.5.22)$$

$$\min_i \left\{ \frac{p_i^t}{p_i^{t'}} \right\} \leq COL^K(u^{t'}, p^t, p^{t'}) \leq COL^L(x^{t'}, p^t, p^{t'}). \qquad (1.5.23)$$

There exists a rather extensive literature that rationalizes cost-of-living indexes that are perhaps more palatable than the Paasche and Laspeyres indices, primarily because they afford more scope for substitution between commodities when relative prices change.

[45] See Pollak (1971) for proofs of these bounds.

For example, the *Fisher "Ideal" Index*,

$$
\begin{aligned}
COL^F(x^t, x^{t'}, p^t, p^{t'}) &= COL^L(x^{t'}, p^t, p^{t'})^{1/2} COL^P(x^t, p^t, p^{t'})^{1/2} \\
&= \left(\frac{p^t \cdot x^{t'}}{p^{t'} \cdot x^{t'}} \frac{p^t \cdot x^t}{p^{t'} \cdot x^t} \right)^{1/2}, \qquad (1.5.24)
\end{aligned}
$$

has been rationalized by a quadratic utility function:

(1.5.24) **Theorem** (Pollak (1971)): The Fisher [1922] "ideal index" is "exact" (*i.e.*, a Konüs "true" index) if and only if the consumer is cost minimizing and the utility function is homothetic quadratic,

$$U(x) = [x' A x]^{1/2}, \qquad (1.5.25)$$

where A is a negative semi-definite symmetric matrix, in which case the cost function is

$$C(u, p) = u(p' A^{-1} p)^{1/2}. \qquad (1.5.26)$$

More recently, Diewert (1976) introduced the notion of a *superlative index number*, one that is exact for a flexible functional form.[46] This conceptualization led to some remarkable theorems rationalizing well-known indexes.

(1.5.27) **Theorem** (Diewert (1976)): If and only if the consumer is cost minimizing and preferences are homothetic and the unit cost function is translog,[47]

$$\ln \Pi(p) = \alpha_0 + \sum_i \alpha_i \ln p_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} \ln p_i \ln p_j, \qquad (1.5.28)$$

the Konüs cost-of-living index is given by

[46]The notion of a flexible functional form—one that can provide a second-order approximation to an arbitrary "true" function at a point—was introduced by Diewert (1971).

[47]Recall that preferences are homothetic if and only if the cost function has the structure, $C(u,p) = \Psi(u)\Pi(p)$, where Ψ is the identity function for some normalization of the utility function. The translog flexible functional form was introduced by Christensen, Jorgenson, and Lau (1973).

$$\begin{aligned}\ln COL^K(u,p^t,p^{t'}) &= \frac{1}{2}\sum_i\left(\frac{p_i^t x_i^t}{p^t\cdot x^t}+\frac{p_i^{t'}x_i^{t'}}{p^{t'}\cdot x^{t'}}\right)(\ln p_i^t-\ln p_i^{t'})\\ &=:\frac{1}{2}\sum_i(s_i^t+s_i^{t'})\ \ln\frac{p_i^t}{p_i^{t'}}\\ &=:\ln COL^T(x^t,x^{t'},p^t,p^{t'}),\end{aligned}\tag{1.5.29}$$

where $COL^T(x^t,x^{t'},p^t,p^{t'})$ is the Törnqvist (1936) cost-of-living index.

Again, the beauty of this result is that a completely data-based cost-of-living index with a proven record of practical usefulness is fully characterized by a particular preference ordering, in that the Törnqvist index is consistent with consumer behavior if and only if preferences are homothetic with a translog unit cost function.

Of course, the problem with the utility rationalization of the Törnqvist price index is that the homotheticity of preferences is an unrealistic assumption. Another result that does not rely on homotheticity is the following:

(1.5.30) **Theorem** (Diewert (1976)): If the consumer is cost minimizing and the cost function is translog,

$$\begin{aligned}\ln C(u,p) &= \alpha_0+\sum_i\alpha_i\ln p_i+\alpha_u\ln u+\frac{1}{2}\sum_i\sum_j\beta_{ij}\ln p_i\ln p_j\\ &\quad+\beta_{uu}(\ln u)^2+\sum_i\beta_{ui}\ln u\ln p_i,\end{aligned}\tag{1.5.31}$$

then

$$COL^K(u^*,p^t,p^{t'})=COL^T(x^t,x^{t'},p^t,p^{t'})\tag{1.5.32}$$

where $u^*=(u^tu^{t'})^{1/2}$.

Thus, a wider class of preferences than that in Theorem 1.5.27 is exact for the Törnqvist index if we specify the geometric mean of the utility levels in the two situations as the reference utility level in the Konüs cost-of-living index. In a sense, then, this result is a flexible-form generalization of the quadratic preference rationalization of the Fisher ideal index (a geometric mean of the Paasche and Laspeyres indexes) in Theorem 1.5.24. Since the Törnquist index is computed using only quantity and price data, the specialized choice of a reference utility level in the Konüs index, while arbitrary, seems harmless enough.[48]

1.5.2 The Malmquist Standard-of-Living Index

The value of a cost-of-living index tells us nothing about the well being of an individual in one situation as compared to another, unless we also have information on expenditure. Moreover, while expenditure information, comparing $COL^K(u^{t'},p^t,p^{t'})$ to the expenditure ratio, $y^t/y^{t'}$, tells us whether the standard of living of a utility-maximizing consumer has increased or decreased, it tells us nothing about the magnitude of the increase or decrease in the standard of living. If utility were cardinally measurable (up to a linear transformation), the utility comparision $U(x^t)/U(x^{t'})$ would be a good measure, but in most economic analyses we are stuck with ordinal utility.

A promising approach is to note that the standard of living is in some conceptual sense dual to the notion of the cost of living and ought to be treated in a symmetric way, just as a quantity index should be symmetrically dual to a price index. This is the brilliant insight of Malmquist (1953).

To pursue this notion, recall that the cost function is a distance function in price space. Now let us show that the Konüs cost-of-living function normalized on base-period utility, $COL^K(u^{t'},p^t,p^{t'})$,

[48]Note that this theorem, unlike Theorem 1.5.27, is not if and only if; that is, assuming that the Konüs index is a Törnqvist index does not imply preferences represented by a translog cost function.

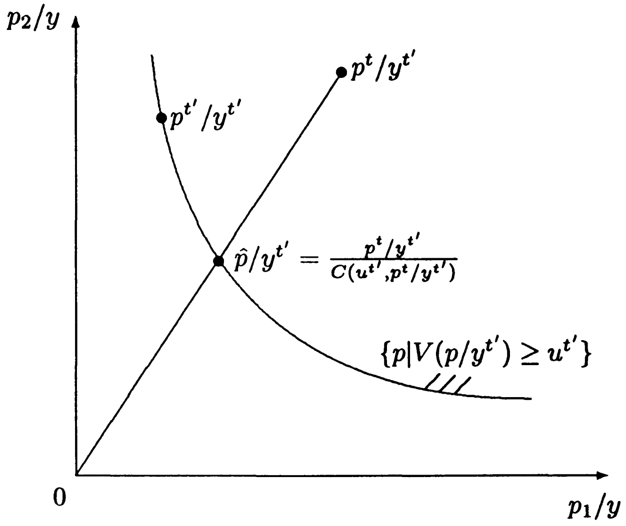

Figure 1.26

is itself a distance function in price space (using normalized prices, p/y, rather than the normalization $y = 1$). Referring to Figure 1.26, we have

$$COL^K(u^{t'},p^t,p^{t'}) = \frac{C(u^{t'},p^t)}{C(u^{t'},p^{t'})} = C(u^{t'},p^t/y^{t'}) \qquad (1.5.33)$$
$$= \frac{\|p^t/y^{t'}\|}{\|\hat{p}/y^{t'}\|} = \frac{\|p^t\|}{\|\hat{p}\|},$$

where the second-to-last identity exploits the distance-function characterization of the consumer expenditure function. Thus, the Konüs cost-of-living index (normalized on base-period utility) is equal to the radial reduction in current-period prices needed to just be able to purchase the utility level of the base period.

This distance-function characterization of the Konüs cost-of-living index suggests that the distance function in quantity space is the natural standard-of-living counterpart. Thus, we introduce the *Malmquist (1953) standard-of-living (quantity) index*:

$$SOL^M(u, x^t, x^{t'}) = \frac{D_I(u, x^t)}{D_I(u, x^{t'})}. \tag{1.5.34}$$

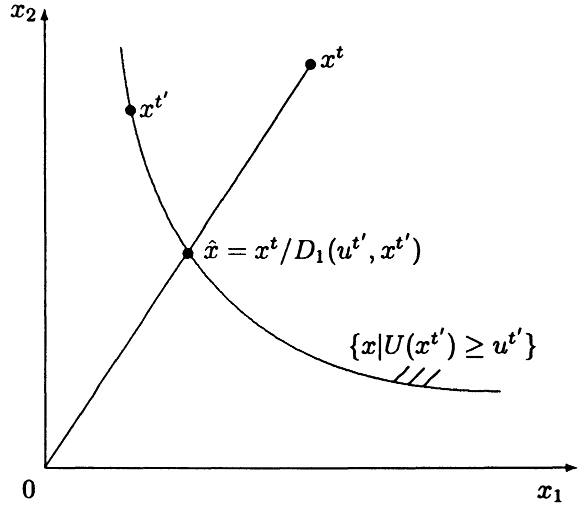

Figure 1.27

This index reflects the standard of living in situation t as compared to situation t' with reference to utility level u, in the sense that it compares the radial contraction (or expansion) in the consumption vector in each situation that is needed to just allow the attainment of utility level u. To make this concept a little more concrete, let us calculate the Malmquist standard-of-living index with reference to base-period utility $u^{t'}$. Thus, in Figure 1.27,

$$SOL^M(u^{t'}, x^t, x^{t'}) = \frac{D_I(u^{t'}, x^t)}{D_I(u^{t'}, x^{t'})} = \frac{D_I(u^{t'}, x^t)}{1} = \frac{\|x^t\|}{\|\hat{x}\|}. \tag{1.5.35}$$

Thus, $SOL^M(u^{t'}, x^t, x^{t'})$ tells us how much the the consumption vector in situation t would have to be contracted to yield utility equal to the situation t' level.

As should be expected, duals to all of the results on cost-of-living indexes hold for standard-of-living indices: First, using the *dual to Shephard's decomposition theorem*,[49]

$$U \text{ is homothetic} \iff D_I(u,x) = \frac{\Lambda(x)}{\Psi(u)}, \tag{1.5.36}$$

where Ψ is increasing and Λ is nondecreasing, concave, and homogeneous of degree one, we have the following:

Exact standard-of-living counterparts to Theorems 1.5.6, 1.5.11, 1.5.18, 1.5.24, and 1.5.27 follow immediately from the symmetric duality between the cost function and the input distance function:

(1.5.39) **Theorem**: $SOL^M(u, x^t, x^{t'})$ is independent of u if and only if the consumer is cost minimizing and preferences are homothetic, in which case

$$SOL^M(u, x^t, x^{t'}) = \frac{\Lambda(x^t)}{\Lambda(x^{t'})}. \tag{1.5.40}$$

(1.5.41) **Theorem**: Laspeyres and Paasche quantity indexes,

$$SOL^M(u^t, x^t, x^{t'}) = \frac{\sum_i p_i^t\, x_i^t}{\sum_i p_i^t\, x_i^{t'}} =: SOL^P(p^t, x^t, x^{t'})$$

and

$$SOL^M(u^{t'}, x^t, x^{t'}) = \frac{\sum_i p_i^{t'}\, x_i^t}{\sum_i p_i^{t'}\, x_i^{t'}} =: SOL^L(p^{t'}, x^t, x^{t'}), \tag{1.5.42}$$

if and only if preferences are represented by

[49] See footnote 44 above.

$$D_I(u,x) = \sum_i a_i(u)x_i, \tag{1.5.43}$$

where the a_i, $i = 1, \ldots, n$, are nonincreasing functions, in which case the cost function is

$$C(u,p) = \min\left\{\frac{p_1}{a_1(u)}, \cdots, \frac{p_n}{a_n(u)}\right\}. \tag{1.5.44}$$

(1.5.45) **Theorem**: The Fisher ideal quantity index,

$$\begin{aligned} SOL^F(p^t, p^{t'}, x^t, x^{t'}) &= SOL^L(p^{t'}, x^t, x^{t'})^{1/2} SOL^P(p^t, x^t, x^{t'})^{1/2} \\ &= \left(\frac{x^t \cdot p^{t'}}{x^{t'} \cdot p^{t'}} \frac{x^t \cdot p^t}{x^{t'} \cdot p^t}\right)^{1/2}, \end{aligned} \tag{1.5.46}$$

is exact if and only if the consumer is cost minimizing and preferences are homothetic quadratic.

(1.5.47) **Theorem**: If and only if the consumer is cost minimizing and preferences are homothetic, in which case $D_I(u,x) = \Psi(u)^{-1}\Lambda(x)$, and the function Λ is homothetic translog, the Malmquist standard-of-living index is given by

$$\begin{aligned} \ln SOL^M(u, x^t, x^{t'}) &= \frac{1}{2}\sum_i \left(\frac{p_i^t x_i^t}{p^t \cdot x^t} + \frac{p_i^{t'} x_i^{t'}}{p^{t'} \cdot x^{t'}}\right)(\ln x_i^t - \ln x_i^{t'}) \\ &=: \frac{1}{2}\sum_i (s_i^t + s_i^{t'}) \ \ln \frac{x_i^t}{x_i^{t'}} \\ &=: \ln SOL^T(p^t, p^{t'}, x^t, x^{t'}), \end{aligned} \tag{1.5.48}$$

where $SOL^T(p^t, p^{t'}, x^t, x^{t'})$ is the Törnqvist (1936) quantity index.

(1.5.49) **Theorem**: If the distance function is translog,

$$\ln D_I(u,x) = \alpha_0 + \sum_i \alpha_i \ln x_i + \alpha_u \ln u + \frac{1}{2}\sum_i \sum_j \beta_{ij} \ln x_i \ln x_j + \beta_{uu}(\ln u)^2 + \sum_i \beta_{ui} \ln u \ln x_i, \tag{1.5.50}$$

then

$$SOL^M(u^*, x^t, x^{t'}) = SOL^T(p^t, p^{t'}, x^t, x^{t'}), \tag{1.5.51}$$

where $u^* = (u^t u^{t'})^{1/2}$.

1.5.3 The Malmquist Standard-of-Living Index and Income Deflation

Increases in the standard of living, or "real income," are commonly calculated by deflating the increase in nominal income by a price, or cost-of-living, index. The theoretical question thus arises of whether deflation by a Konüs cost-of-living index is related to the Malmquist standard-of-living index. The short answer is no—unless very strong (and unrealistic) restrictions on preferences are satisfied. The special case in which such a relationship exists is the case where p^t and x^t are conjugate pairs relative to $u^{t'}$ (see Section 1.4.1 above). This case is illustrated in Figure 1.28. In this example (which presupposes cost minimization), p^t supports $L(u^{t'})$ at $x^t/D_I(u^{t'}, x^t)$; consequently,

$$\begin{aligned} D_I(u^{t'}, x^t) &= \frac{p^t \cdot x^t}{C(u^{t'}, p^t)} \\ &= \frac{p^t \cdot x^t \;/\; p^{t'} \cdot x^{t'}}{C(u^{t'}, p^t) \;/\; C(u^{t'}, p^{t'})}, \end{aligned} \tag{1.5.52}$$

or

$$SOL^M(u^{t'}, x^{t'}, x^t) = \frac{y^t \;/\; y^{t'}}{COL^K(u^{t'}, p^t, p^{t'})}. \qquad (1.5.53)$$

That is, the Malmquist index of the increase in the standard of living between situations t' and t equals the ratio of the nominal incomes in the two situations deflated by the corresponding Konüs cost-of-living index.

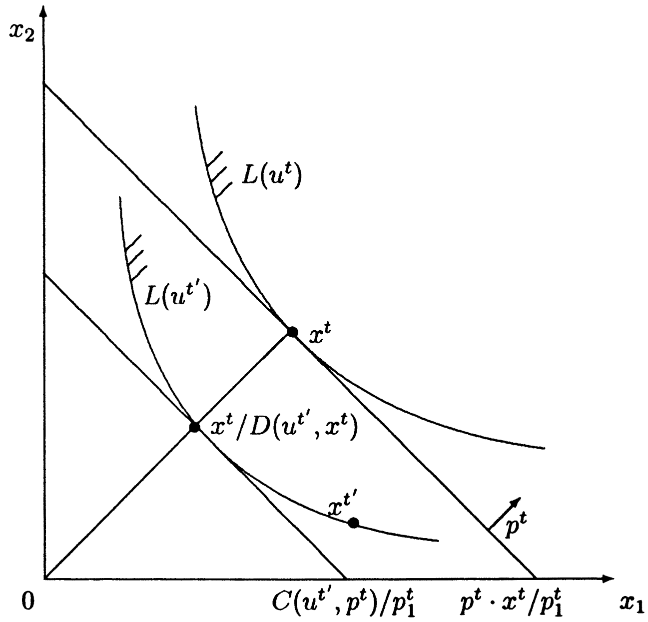

Figure 1.28

The following result is now obvious:

(1.5.54) **Theorem**: The equivalence of Konüs-based income deflation and the Malmquist measurement of changes in the standard of living, (1.5.53), holds for all possible situations if and only if preferences are homothetic.

1.5.4 Malmquist Productivity Indexes

While the Malmquist quantity index was formulated as a standard-of-living index in the theory of the consumer, it is clearly applicable to production theory as well. In Section 1.3, the output distance function was promoted as a natural generalization of the single-output measure of efficiency given by the ratio of actual to potential (frontier) output: $u/U(x)$. An alternative measure of efficiency is the inverse of the input distance function. For output/input combinations in the technology set, all of these indexes lie in the $(0, 1]$ interval. These efficiency indexes take the technology as given and aggregate across output or input quantities to obtain a scalar measure of the shortfall from the (weakly or strongly) efficient production frontier.

The distance functions can alternatively be employed as indexes of technological change, or differences in technologies across production units, assuming that production units operate efficiently. Thus, efficiency measurement takes the technology as given and attempts to obtain a scalar index of the distance from the frontier, while productivity measurement takes efficiency as given and attempts to obtain a scalar index of the change in (or difference in) technologies.[50] But there is a close parallel between these two measurement problems. The classic study by Solow (1957) measured technological change for single-output technologies by the ratio,

$$PROD^{t'}(x^t) := \frac{U^t(x^t)}{U^{t'}(x^t)} = \frac{u^t}{U^{t'}(x^t)}, \tag{1.5.55}$$

where U^t and $U^{t'}$ are the technologies in the current period and the base period (more generally, in situation t and situation t'), respectively, and the second identity holds only if the producer is operating (technically) efficiently in period t. Thus, Solow's index of technological change is the ratio of potential output in the current period (or actual output if the firm is operating technically

[50] In fact, these two measurement concepts can be combined; see the paper by Färe, Grosskopf, and Roos in this volume.

efficiently) to potential output in the base period t', using the input vector of period t, and is greater than one for technological progress (and less than one for technological regress). This index uses the current-period input quantities x^t as a point of reference.

One can, alternatively, characterize Solow's index of technological change using the base-period input quantities as a reference:

$$PROD^t(x^{t'}) := \frac{U^t(x^{t'})}{U^{t'}(x^{t'})} = \frac{U^t(x^{t'})}{u^{t'}}, \tag{1.5.56}$$

where the second identity holds only if the firm is operating efficiently in period t'. This index is the ratio of potential output in period t to potential output in period t' using the input vector of period t' and is greater than one for technological progress. Of course, Solow assumed Hicks neutral technological change, $U^t(x) = A(t)U(x)$, in which case both (1.5.55) and (1.5.56) simplify to the ratio $A(t)/A(t')$.

Caves, Christensen, and Diewert (1982) generalized the Solow notion of technological change (or comparison) to the case of multiple outputs using distance functions. The index (1.5.55), when greater than 1, tells us how much the scalar output in situation t would have to be contracted (dividing by $PROD^{t'}(x^t)$), using input vector x^t, if it were stuck with the situation t' technology. The index (7.5.56), when greater than 1, tells us how much the scalar output in situation t' could be expanded (multiplying by $PROD^t(x^{t'})$), using input vector $x^{t'}$, if it could produce with the situation t technology. These notions of contractions or expansions of output under alternative technologies can be generalized quite naturally to multiple outputs by using the radial contractions and expansions implicit in output distance functions corresponding to alternative technologies. Thus, corresponding to (1.5.55) and (1.5.56), we have, for multiple outputs, the *output-based, technology-t'-based productivity index*:

$$PROD_O^{t'}(u^t, x^t) = \frac{D_O^{t'}(u^t, x^t)}{D_O^{t'}(u^{t'}, x^{t'})} = D_O^{t'}(u^t, x^t); \tag{1.5.57}$$

and the *output-based, technology-t-based productivity index*:

$$PROD_O^t(u^{t'}, x^{t'}) = \frac{D_O^t(u^t, x^t)}{D_O^t(u^{t'}, x^{t'})} = \frac{1}{D_O^t(u^{t'}, x^{t'})}, \tag{1.5.58}$$

where D_O^τ is the output distance function that characterizes the situation-τ technology and where the second identity in (1.5.57) and (1.5.58) hold only if the firm is operating efficiently in period t' and period t, respectively.

In the case of temporal productivity comparisons with technological progress, $D_O^{t'}(u^t, x^t) > 1$ and $D_O^t(u^{t'}, x^{t'}) < 1$, in which case (1.5.57) tells us the minimal radial reduction in current-period output quantities required to contract the output vector to the base-period production possibility set using the current-period input vector (recall that $u^t \in P^{t'}(x^t)$ if and only if $D_O^{t'}(u^t, x^t) \le 1$) and (1.5.58) tells us the maximal radial increase in base-period output quantities that would be possible with the current-period technology using the base-period input vector ($u^{t'} \in P^t(x^{t'})$ if and only if $D_O^t(u^{t'}, x^{t'}) \le 1$). In Figure 1.29, $PROD_O^{t'}(u^t, x^t) = \|u^t\| \,/\, \|\hat{u}\|$ and $PROD_O^t(u^{t'}, x^{t'}) = \|\bar{u}\| \,/\, \|u^{t'}\|$.

Of course, it is just as natural to measure productivity changes in input space—that is, in terms of radial contractions possible, or radial expansions required, of situation-τ input vectors to produce the situation-τ output vector under an alternative technology. Thus, dual to (1.5.57) and (1.5.58), we have the *input-based, technology-t'-based productivity index*:

$$PROD_I^{t'}(u^t, x^t) = \frac{D_I^{t'}(u^{t'}, x^{t'})}{D_I^{t'}(u^t, x^t)} = \frac{1}{D_I^{t'}(u^t, x^t)}; \tag{1.5.59}$$

and the *input-based, technology t-based productivity index*:

$$PROD_I^t(u^{t'} x^{t'}) = \frac{D_I^t(u^{t'}, x^{t'})}{D_I^t(u^t, x^t)} = D_I^t(u^{t'}, x^{t'}), \tag{1.5.60}$$

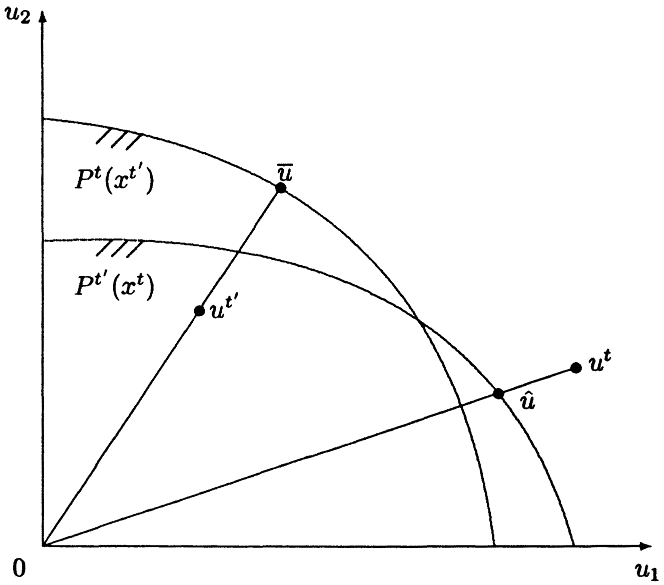

Figure 1.29

where the second identities in (1.5.59) and (1.5.60) hold only if the firm is operating efficiently in period t and period t', respectively. For explication purposes, assume again temporal technological progress, $D_I^{t'}(u^t, x^t) < 1$ and $D_I^t(u^{t'}, x^{t'}) > 1$, in which case (1.5.59) tells us the minimal radial increase in current-period input quantities required to expand the input vector to the base-period input requirement set corresponding to the current-period output vector ($x^t \in L^{t'}(u^t)$ if and only if $D_I^{t'}(u^t, x^t) \geq 1$) and (1.5.60) tells us the radial decrease in base-period input quantities that would be possible with the current-period technology to produce the base-period output vector ($x^{t'} \in L^t(u^{t'})$ if and only if $D_I^t(u^{t'}, x^{t'}) \geq 1$). In Figure 1.30, $PROD_I^{t'}(u^t, x^t) = \|\hat{x}\| \;/\; \|x^t\|$ and $PROD_I^t(u^{t'}, x^{t'}) = \|x^{t'}\| \;/\; \|\bar{x}\|$.

Superlative productivity index number theorems, analogous to those for cost-of-living and standard-of-living indices in Section 1.5.1 and 1.5.2, but requiring that second-order terms be identical for the

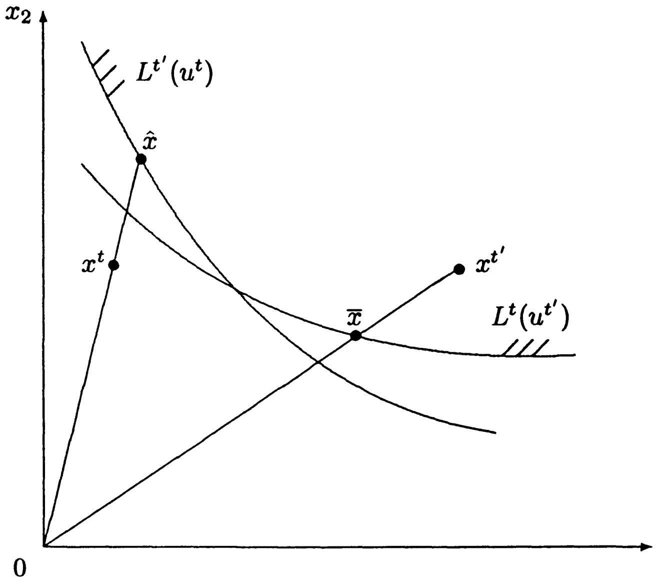

Figure 1.30

two technologies, were proved by Caves, Christensen, and Diewert (1982) (*e.g.*, in this case a Törnqvist index is exact for a translog input or output distance function, with the required conditions for second-order terms).[51]

1.5.5 Other Indices Based on the Distance Function

The distance function has arisen as a natural construct in other index-number problems. For example, Atkinson's (1970) inequality index, and the refinements and extensions to poverty measurement of Blackorby and Donaldson (1978, 1980a, 1980b), are distance functions applied to social evaluation functions defined on the space

[51]See the contribution to this volume by Färe, Grosskopf, and Roos for much more on Malmquist productivity indices.

of income distributions. Also, the notion of household equivalence scales, deflating the household consumption vector by the number of "adult equivalents," can be interpreted as a distance function (Muellbauer (1977)). See Deaton (1978) for a brief discussion of these concepts.

1.6 CONCLUDING REMARKS

In this paper, I have argued that the distance function is

- a natural functional representation of a multiple-output production technology,
- a natural measure of technical (in)efficiency,
- a natural dual to the commonly employed cost function,
- a natural construct for examining the comparative statics of shadow-price (inverse Hicksian) demand functions,

and

- a natural construct for calculating standard-of-living indices that are dual to the Könus cost-of-living index, or productivity indices.

It is a tribute to Sten Malmquist that his prescient 1953 paper has had such a profound effect on the development and application of modern consumer and producer theory. It took half a century for Hotelling's (1932) penetrating use of the duality between the expenditure function and the preferences to find its way into economic theory textbooks; almost half a century after Malmquist's equally penetrating paper, it is high time that the distance function takes its rightful position next to the expenditure function in modern textbooks.

References

Atkinson, A. B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory* 2, 244–263.

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

Blackorby, C., and D. Donaldson (1978), "Measures of Inequality and Their Meaning in Terms of Social Welfare," *Journal of Economic Theory* 18, 59–80.

Blackorby, C., and D. Donaldson (1980a), "A Theoretial Treatment of Indices of Absolute Inequality," *International Economic Review* 21, 107–136.

Blackorby, C., and D. Donaldson (1980b), "Ethical Indices for the Measurement of Poverty," *Econometrica* 48, 1053–1060.

Blackorby, C., and R. R. Russell (1976), "Functional Structure and the Allen Elasticities of Substitution: An Application of Duality Theory," *Review of Economic Studies* 43, 285–292.

Blackorby, C., and R. R. Russell (1981), "The Morishima Elasticity of Substitution: Symmetry, Constancy, Separability, and Its Relationship to the Hicks and Allen Elasticities," *Review of Economic Studies* 48, 147–158.

Blackorby, C., and R. R. Russell (1996), "Two-Stage Budgeting: An Extension of Gorman's Theorem," *Economic Theory* 9, 185-193.

Bol, G. (1986), "On Technical Efficiency Measures: A Remark," *Journal of Economic Theory* 38, 380–385.

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 Exact Input, Output, Productivity, and Technical Change Measurement," Working Paper 96–05, Maryland Agricultural Experiment Station, 1996.

Chambers, R. G., Y. Chung, and R. Färe (1996), "Benefit and Distance Functions," *Journal of Economic Theory* 70, 407-419.

Charnes, A., W. W. Cooper, and E. Rhodes (1978), "Measuring Efficiency of Decision Making Units," *European Journal of Operational Research*, 6, 429–444.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1973), "Transcendental Logarithmic Production Frontiers," *Review of Economics ans Statistics* 65, 28–45.

Deaton, A. (1978), "The Distance Function in Consumer Behaviour with Applications to Index Numbers and Optimal Taxation," *Review of Economic Studies* 45, 391–405.

Deaton, A. (1981), "Optimal Taxes and the Structure of Preferences," *Econometrica* 49, 1245–1260.

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

Diewert, W. E. (1971), "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy* 79, 481-507.

Diewert, W. E. (1974), "Applications of Duality Theory," in *Frontiers of Quantitative Economics*, Vol. 2, edited by M. Intriligator and D. Kendrick, Amsterdam: North-Holland.

Diewert, W. E. (1976) "Exact and Superlative Index Numbers," *Journal of Econometrics* 4, 115–145.

Dmitruk, V., and G. A. Koshevoy (1991), "On the Existence of a Technical Efficiency Criterion," *Journal of Economic Theory* 55, 121–144.

Eichhorn, W., and J. Voeller (1976), *Theory of the Price Index: Fisher's Test Approach and Generalisations*, Lecture Notes in Economics and Mathematics, Vol. 140, Berlin: Springer-Verlag.

Färe, R., and S. Grosskopf (1990), "A Distance Function Approach to Price Efficiency," *Journal of Political Economy* 43, 123–126.

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

Färe, R., S. Grosskopf, and J. Nelson (1990), "On Price Efficiency," *International Economic Review* 31, 709–720.

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

Färe, R., C. A. K. Lovell, and K. Zieschang (1983), "Measuring the Technical Efficiency of Multiple Output Production Technologies," in *Quantitative Studies in Production and Prices*, edited by W. Eichhorn R. Henn, O. Opitz, and R. W. Shephard, Würzburg: Physica-Verlag.

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.

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

Fuss, M., and D. McFadden (1978), *Production Economics; A Dual Approach to Theory and Applications*, Amsterdam: North-Holland.

Gorman, W. M. (1976), "Tricks with Utility Functions," in *Essays in Economic Analysis*, edited by M. Artis and R. Nobay, Cambridge: Cambridge University Press.

Hackman, S., and R. R. Russell (1995), "Continuity and Duality," *Journal of Productivity Analysis* 6, 99–116.

Hanoch, G. (1978), "Symmetric Duality and Polar Production Functions, in Fuss and McFadden (1978).

Hotelling, H. (1932), "Edgeworth's Taxation Paradox and the Nature of Supply and Demand Functions," *Journal of Political Economy* 40, 557–616.

Houthakker, H. S. (1961), "The Present State of Consumption Theory." *Econometrica* 29, 704–740.

Konüs, A. A. (1924), "The Problem of the True Index of the Cost of Living," *Econometrica* 7, 10–29 (translation).

Koopmans, T. J. (1951), "Analysis of Production as an Efficient Combination of Activities," Chapter 3 in *Activity Analysis of Production and Allocation*, edited by T. C. Koopmans, New York: Wiley.

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

Luenberger, D. G. (1995), *Microeconomic Theory*, New York: McGraw Hill.

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

McFadden, D. (1978), "Cost, Revenue, and Profit Functions," in Fuss, M., and D. McFadden (1978).

Muellbauer, J. (1977), "Testing the Barten Model of Household Composition Effects and the Cost of Children," *Economic Journal* 87, 460–487.

Pollak, R. A. (1971), "The Theory of the Cost of Living Index," Research Discussion Paper, Office of Prices and Living Conditions, Bureau of Labor Statistics, Washington, D.C. In *Price Level Measurement: Proceedings from a Conference Sponsered by Statistics Canada*, edited by W.E. Diewert and C. Montmarquette, Ottawa: Statistics Canada, 1983. Reprinted in *Price Level Measurement*, Amsterdam: North-Holland, edited by W. E. Diewert, 1990, and in R. A. Pollak, *The Theory of the Cost of Living Index*, Oxford: Oxford University Press, 1989.

Primont, D., and C. Sawyer (1993), "Recovering the Production Technology from the Cost Function," *Journal of Productivity Analysis* 4, 347–352.

Russell, R. R. (1985), "Measures of Technical Efficiency," *Journal of Economic Theory* 35, 109–126.

Russell, R. R. (1987), "On the Axiomatic Approach to the Measurement of Technical Efficiency," in *Measurement in Economics: Theory and Applications of Economic Indices*, edited by W. Eichhorn, Heidelberg: Physica-Verlag.

Russell, R. R. (1990), "Continuity of Measures of Technical Efficiency," *Journal of Economic Theory* 51, 255–267.

Samuelson, P. A. (1947), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Samuelson, P. A. (1950), "The Problem of Integrability in Utility Theory," *Economica* 17, 335–385.

Samuelson, P. A. (1960), "Structure of a Minimum Equilibrium System," in *Essays in Economics and Econometrics*, edited by R.

W. Pfouts, Chapel Hill: University of North Carolina Press.

Seiford, L. M., (1996) "A Bibliography of Data Envelopment Analysis (1978–1990)," in *Proceedings of a Conference on New Uses of DEA*, Quorum Books.

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

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

Solow, R. M. (1957), "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39, 312–320.

Törnqvist, L. (1936), "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin*, No. 10, 1–8.

Varian, H. R. (1992), *Microeconomic Analysis*, New York: Norton.

Zieschang, K. D. (1984), "An Extended Farrell Technical Efficiency Measure," *Journal of Economic Theory* 33, 387–396.

2

INPUT PRICE, QUANTITY, AND PRODUCTIVITY INDEXES FOR A REVENUE-CONSTRAINED FIRM

by

Bert M. Balk*

2.1 INTRODUCTION

This paper develops the micro-economic theory of input price and quantity indexes and input-based productivity indices for a revenue-constrained firm. Generally speaking, a firm transforms inputs into outputs. When priced, outputs generate revenue and inputs incur cost. It is customary to (partially) model the economic behavior of a firm as cost minimization. Then the input price index is calculated as the ratio of minimum costs under two different price regimes. However, in doing so, one has to condition on certain output variables. One route is to take the output quantities as the conditioning variables. The firm's objective is then conceived as the production of a vector of output quantities with minimal cost. This leads to a theory of input price indexes which is, except for the dimension of the vector of conditioning variables, isomorphic to the theory of the cost-of-living index for consumers.

*The author thanks Robert G. Chambers and R. Robert Russell for their comments on a previous version. The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

Fisher (1995) considered this approach, except for the case of a single- output firm, as being too narrow. According to his view "the demand conditions that a competitive firm takes as given are not represented by a vector of fixed output quantities but by a vector of fixed output prices at which the firm can sell." Thus a more realistic route is to condition on output prices, and to take the firm's objective as the attainment of a target revenue with minimal input cost. This leads to the theory of the revenue-constrained firm, as developed by Färe and Grosskopf (1994). The applicability of this theory is, however, not restricted to single firms. Sidestepping the aggregation issue, Fisher (1995) uses the revenue-constrained firm as a model for a small, fully open economy which trades outputs on world markets at fixed prices.

Having pointed out the significance of the present topic, we proceed to describe the plan of this paper. Section 2 briefly reviews a number of concepts from duality theory. We limit ourselves to those concepts that play a role in the ensuing sections. Section 3 is devoted to the indirect distance function and the indirect cost function. Both sections build heavily on Färe and Primont (1995). Using the concepts introduced in Section 3, Section 4 formally defines the indirect input price and quantity indexes. Their properties will be shown, and we discuss under which assumptions the theoretical indexes can be approximated by or calculated as Fisher or Törnqvist index numbers. Section 5 turns to the definition of productivity indexes, and shows under which assumptions the theoretical productivity indexes transform into compositions of statistical index numbers. By way of conclusion, Section 6 summarizes the main results in non-technical language.

2.2 PRIMAL AND DUAL DIRECT REPRESENTATIONS OF THE TECHNOLOGY

Formally, a firm is an entity transforming inputs into outputs. The input quantities will be represented by an $N-$dimensional vector of non-negative real values $x \equiv (x_1, ..., x_N) \in \Re^N_+$. The output quantities will be represented by an $M-$dimensional vector of non-negative real values $y \equiv (y_1,...,y_M) \in \Re^M_+$. Vectors without superscripts, but occasionally with primes, are variables. In this Chapter we will consider a single firm, which is observed during a number of discrete time periods $0, 1, 2, \ldots$. We let vectors with superscripts represent observations, thus for instance (x^t,y^t) denotes the input and output of the firm in period t. When two time periods are compared, the earlier period (usually $t = 0$) will be called the 'base period' and the later period (usually $t = 1$) will be called the 'comparison period'.

We assume that this firm has access to a certain technology. This technology can change over time. The technology in period t is given by the set of all feasible input-output combinations

$$S^t \equiv \{(x, y) \mid x \text{ can produce } y \text{ in period } t\}. \tag{2.2.1}$$

Alternatively, the period t technology can be described by the input sets

$$L^t(y) \equiv \{x \mid (x, y) \in S^t\} \text{ for all } y, \tag{2.2.2}$$

or by the output sets

$$P^t(x) \equiv \{y \mid (x, y) \in S^t\} \text{ for all } x. \tag{2.2.3}$$

Following Färe and Primont (1995) we assume that the technology satisfies the following axioms:[1]

P.1 $0_M \in P^t(x)$ for all x (inactivity is possible).

[1]Notation: 0_M is a vector of M zero's and $y' \leq y$ means that $y'_m \leq y_m (m = 1, ..., M)$.

P.2S If $y \in P^t(x)$ and $y' \leq y$ then $y' \in P^t(x)$ (strong disposability of outputs).

P.3 $P^t(x)$ is bounded for all x (scarcity).

P.4 $P^t(x)$ is closed for all x.

P.5 If $y \neq 0_M$ then $y \notin P^t(0_N)$ (no free lunch).

P.6S If $y \in P^t(x)$ and $x' \geq x$ then $y \in P^t(x')$ (strong disposability of inputs).

P.7 $L^t(y)$ is closed for all y.

P.8 $L^t(y)$ is convex for all y.

P.9 $P^t(x)$ is convex for all x.

An important representation of the period t technology is provided by the so-called (direct) *input distance function*

$$D_i^t(x,y) \equiv \sup\{\delta \mid \delta > 0,\ x/\delta \in L^t(y)\}. \tag{2.2.4}$$

This function was introduced by Shephard (1953) and independently (in the consumer context) by Malmquist (1953). Since the inputs are disposable we have the following equivalence

$$x \in L^t(y) \text{ if and only if } D_i^t(x,y) \geq 1. \tag{2.2.5}$$

The input distance function is linearly homogeneous in x and nonincreasing in y.

Similarly, the (direct) *output distance function* is defined by

$$D_o^t(x,y) \equiv \inf\{\delta \mid \delta > 0,\ y/\delta \in P^t(x)\}. \tag{2.2.6}$$

Since the outputs are disposable we have the following equivalence

$$y \in P^t(x) \text{ if and only if } D_o^t(x,y) \leq 1. \tag{2.2.7}$$

The output distance function is nonincreasing in x and linearly homogeneous in y. The output distance function measure of *local* *scale elasticity* reflects the sensitivity of the output distance function with respect to changes in the input quantity vector. Provided that the output distance function is continuously differentiable, local scale elasticity is defined by

$$\epsilon_o^t(x,y) \equiv -x\nabla_x D_o^t(x,y)/D_o^t(x,y), \tag{2.2.8}$$

where ∇_x denotes the vector of first order derivatives with respect to x, and $x\nabla_x D_o^t(x,y)$ is the inner product of x and $\nabla_x D_o^t(x,y)$. Notice that, since $D_o^t(x,y)$ is linearly homogeneous in y, $\epsilon_o^t(x,\lambda y) = \epsilon_o^t(x,y)$ for $\lambda > 0$.

The period t technology exhibits *input homotheticity* if for all y $L^t(y) = H^t(y)L^t(1_M)$, where $H : \Re_+^M \to \Re_+$ is a nondecreasing function consistent with the axioms and 1_M is a vector of M ones. Thus each input set $L^t(y)$ is a radial expansion or contraction of $L^t(1_M)$. Input homotheticity is easily shown to be equivalent to

$$D_i^t(x,y) = D_i^t(x,1_M)/H^t(y). \tag{2.2.9}$$

If the technology exhibits input homotheticity then for all y $L^t(\theta y) = (H^t(\theta y)/H^t(y))L^t(y)$ $(\theta > 0)$. If the function $H^t(y)$ is linearly homogeneous, then for all y $L^t(\theta y) = \theta L^t(y)$ $(\theta > 0)$. In this case we say that the technology exhibits global *constant returns to scale* (CRS). Equivalent conditions for global CRS are: $P^t(\theta x) = \theta P^t(x)$ $(\theta > 0)$ for all x; $S^t = \theta S^t$ $(\theta > 0)$; $D_i^t(x,y)$ is homogeneous of degree -1 in y; $D_o^t(x,y)$ is homogeneous of degree -1 in x. A final characterization of global CRS follows from (2.2.8), namely

$$\epsilon_o^t(x,y) = 1 \text{ for all } (x,y). \tag{2.2.10}$$

Let $w \in \Re_{++}^N$ be a vector of positive input prices. The (direct) *cost function* is, for all producible y, defined by

$$\begin{aligned} C^t(w,y) &\equiv \min_x\{wx \mid (x,y) \in S^t\} \qquad (2.2.11)\\ &= \min_x\{wx \mid D_i^t(x,y) \geq 1\}\\ &= \min_x\{wx/D_i^t(x,y)\}, \end{aligned}$$

using (2.2.2), (2.2.5) and Lemma (3.A.5) of Färe and Primont (1995). Recall that wx denotes the inner product of w and x, thus $wx \equiv \sum_{n=1}^{N} w_n x_n$. $C^t(w,y)$ is the minimum cost of producing output y with period t's technology when the input prices are w. The cost function is linearly homogeneous in w and nondecreasing in y. It is easy to show that input homotheticity is equivalent to

$$C^t(w,y) = H^t(y)C^t(w,1_M). \tag{2.2.12}$$

The equivalence rests on Färe and Primont's (1995) duality (II). The period t technology exhibits global CRS if and only if $C^t(w,y)$ is linearly homogeneous in y.

Based on the cost function the (direct) *input price index* is defined by

$$P_i^t(w,w',y) \equiv C^t(w,y)/C^t(w',y), \tag{2.2.13}$$

where w and w' are two input price vectors. This index was proposed by Court and Lewis (1942-43). It is analogous to the Konüs (1924) cost-of-living index in the consumer context. The input price index takes the output quantity vector y as given and compares the minimum cost of producing y under two input price vectors w, w' using the period t technology. In fact, (2.2.13) defines a family of input price indexes, namely for different values of t and y. It is important to know under which circumstances the dependency on y vanishes. The following result is well-known and its proof is given for completeness' sake.

(2.2.14) **Proposition**: $P_i^t(w,w',y)$ is independent of y if and only if the period t technology exhibits input homotheticity.

Proof: By (2.2.12), input homotheticity implies immediately that $P_i^t(w,w',y) = C^t(w,1_M)/C^t(w',1_M)$, independent of y. Reversely, assume that $P_i^t(w,w',y)$ is independent of y. This implies that $C^t(w,y)/C^t(w',y) = g^t(w)/g^t(w')$ for some function $g^t(w)$, or

$$C^t(w,y) = g^t(w)C^t(w',y)/g^t(w').$$

The lefthand side of this equation is independent of w'. Thus the righthand side must also be independent of w', which implies that there exists a function $\phi^t(y)$ such that

$$C^t(w,y) = g^t(w)\phi^t(y) \text{ for all } w, y.$$

In particular $C^t(w, 1_M) = g^t(w)\phi^t(1_M)$. Combining this with the foregoing result we obtain

$$C^t(w,y) = C^t(w,1_M)H^t(y),$$

where $H^t(y) \equiv \phi^t(y)/\phi^t(1_M)$. But this is equivalent to input homotheticity.

Q.E.D.

The (direct) *input quantity index* can be based on the input distance function and is therefore often called the Malmquist input quantity index. Thus we define

$$Q_i^t(x,x',y) \equiv D_i^t(x,y)/D_i^t(x',y), \tag{2.2.15}$$

where x and x' are the two input quantity vectors compared. In fact, (2.2.15) defines a family of input quantity indices, namely for various values of t and y, and the following well-known result is important.

(2.2.16) **Proposition**: $Q_i^t(x,x',y)$ is independent of y if and only if the period t technology exhibits input homotheticity.

Proof: (Färe and Primont 1995) Using (2.2.9), the proof is virtually identical to that of Proposition 2.2.14.

Let $p \in \Re^M_{++}$ be a vector of positive output prices. The (direct) *revenue function* (sometimes called restricted profit function) is defined by

$$R^t(x,p) \equiv \max_y\{py \mid (x,y) \in S^t\} \tag{2.2.17}$$

$$\begin{aligned} &= \max_y \{py \mid D_o^t(x,y) \le 1\} \\ &= \max_y \{py / D_o^t(x,y)\}, \end{aligned}$$

using (2.2.3), (2.2.7) and Lemma (3.A.7) of Färe and Primont (1995). $R^t(x,p)$ is the maximum revenue that can be obtained by using input x with period $t's$ technology when the output prices are p. The revenue function is linearly homogeneous in p and nondecreasing in x. The period t technology exhibits global CRS if and only if $R^t(x,p)$ is linearly homogeneous in x.

The sensitivity of the revenue function with respect to changes in x can be related to the output distance function measure of local scale elasticity. We firstly observe that the final line of (2.2.17) is equivalent to

$$\ln R^t(x,p) = \max_y \{\ln(py) - \ln D_o^t(x,y)\}. \tag{2.2.18}$$

If the revenue function is continuously differentiable, applying the Envelope Theorem to (2.2.18) yields

$$\partial \ln R^t(x,p)/\partial x_n = -\partial \ln D_o^t(x,y^*)/\partial x_n \ (n = 1, ..., N), \tag{2.2.19}$$

where y^* is the solution to the revenue maximization problem. Multiplying the lefthand side and the righthand side by x_n and summing, we obtain

$$\begin{aligned} \epsilon_R^t(x,p) &\equiv x\nabla_x R^t(x,p)/R^t(x,p) \qquad (2.2.20) \\ &= -x\nabla_x D_o^t(x,y^*)/D_o^t(x,y^*) = \epsilon_o^t(x,y^*). \end{aligned}$$

In other words, at the optimum point the output distance function measure of local scale elasticity is equal to the elasticity of the revenue function with respect to input.

2.3 THE INDIRECT INPUT DISTANCE FUNCTION, COST FUNCTION, AND EFFICIENCY MEASURES

After all this groundwork we now consider the indirect, revenue constrained input sets

$$IL^t(p/r) \equiv \{x \mid x \in L^t(y),\ py \geq r\}, \tag{2.3.1}$$

where $r \in \Re_{++}$. $IL^t(p/r)$ is the set of all input quantity vectors x that can 'produce' at least revenue r in period t. Notice that the constraint $py \geq r$ can be rewritten as $(p/r)y \geq 1$. Thus the relevant variable is p/r rather than (p, r). The relation between direct and indirect input sets becomes more transparent if we rewrite (2.3.1) as

$$IL^t(p/r) = \cup_{\{py \geq r\}} L^t(y). \tag{2.3.2}$$

Thus each indirect input set is the union of a number of direct input sets. We assume that the indirect input sets are non-empty, closed and convex. It is easy to see that an alternative representation of the indirect input sets is

$$IL^t(p/r) = \{x \mid R^t(x, p) \geq r\}, \tag{2.3.3}$$

where $R^t(x, p)$ is the (direct) revenue function defined by (2.2.17). Analogous to the definition of the direct input distance function (2.2.4), the *indirect input distance function* is defined by

$$ID_i^t(x, p/r) \equiv \sup\{\delta \mid \delta > 0,\ x/\delta \in IL^t(p/r)\}. \tag{2.3.4}$$

This function was introduced by Shephard (1974). Using the disposability property of the inputs, we obtain the following equivalence

$$x \in IL^t(p/r) \text{ if and only if } ID_i^t(x, p/r) \geq 1. \tag{2.3.5}$$

The indirect input distance function is linearly homogeneous in x and nondecreasing in p/r. Two alternative representations of the indirect distance function will prove useful. Using (2.3.1) and (2.2.5) respectively, we see that

$$
\begin{aligned}
ID_i^t(x,p/r) &= \sup_{\delta,y}\{\delta \mid \delta > 0, x/\delta \in L^t(y), py \geq r\} \qquad (2.3.6)\\
&= \sup_{\delta,y}\{\delta \mid \delta > 0, D_i^t(x/\delta, y) \geq 1, py \geq r\}\\
&= \sup_{\delta,y}\{\delta \mid \delta > 0, D_i^t(x,y) \geq \delta, py \geq r\}\\
&= \sup_{y}\{D_i^t(x,y) \mid py \geq r\}.
\end{aligned}
$$

The final equality reflects the relation of direct and indirect input sets given by (2.3.2). The other representation is obtained by substituting (2.3.3) into (2.3.4), giving

$$
ID_i^t(x,p/r) = \sup\{\delta \mid \delta > 0, R^t(x/\delta, p) \geq r\}. \qquad (2.3.7)
$$

This expression can be used to prove that the period t technology exhibits global CRS if and only if

$$
ID_i^t(x,p/r) = R^t(x,p/r) = R^t(x,p)/r \qquad (2.3.8)
$$

(see Färe and Primont 1995, (4.1.13)). Input homotheticity was introduced in Section 2. Its implication for the structure of the indirect input distance function is the topic of the next Proposition.

(2.3.9) **Proposition**: The period t technology exhibits input homotheticity if and only if $ID_i^t(x,p/r) = D_i^t(x,1_M)/IH^t(p/r)$ for some nonincreasing function $IH^t(p/r)$.

Proof: Recall (2.2.9) for the definition of input homotheticity, and substitute into the last line of (2.3.6). Then

$$
\begin{aligned}
ID_i^t(x,p/r) &= \sup_{y}\{D_i^t(x,1_M)/H^t(y) \mid py \geq r\}\\
&= D_i^t(x,1_M)/\inf_{y}\{H^t(y) \mid py \geq r\}\\
&= D_i^t(x,1_M)/IH^t(p/r),
\end{aligned}
$$

where $IH^t(p/r)$ is defined by the last equality. The reverse direction uses duality (V) of Färe and Primont (1995):

$$\begin{aligned} D_i^t(x,y) &= \inf_{p/r}\{ID_i^t(x,p/r) \mid py \geq r\} \\ &= \inf_{p/r}\{D_i^t(x,1_M)/IH^t(p/r) \mid py \geq r\} \\ &= D_i^t(x,1_M)/\sup_{p/r}\{IH^t(p/r) \mid py \geq r\} \\ &= D_i^t(x,1_M)/H^t(y), \end{aligned}$$

where $H^t(y)$ is defined by the last equality. This is equivalent to input homotheticity.

Q.E.D.

Analogous to the definition of the direct cost function (2.2.11), the *indirect (revenue constrained) cost function* is defined by

$$\begin{aligned} IC^t(w,p/r) &\equiv \min_x\{wx \mid x \in IL^t(p/r)\} \qquad (2.3.10) \\ &= \min_x\{wx \mid ID_i^t(x,p/r) \geq 1\} \\ &= \min_x\{wx/ID_i^t(x,p/r)\} \\ &= \min_y\{C^t(w,y) \mid py \geq r\} \\ &= \min_x\{wx \mid R^t(x,p) \geq r\}. \end{aligned}$$

The first equality was obtained by using (2.3.5), the second by using Lemma (3.A.5) of Färe and Primont (1995), the third by using (2.3.1) and (2.2.11) respectively, and the fourth by using (2.3.3). The indirect cost function gives the minimum cost of 'producing' revenue r with period t's technology when the input prices are w and the output prices are p. The cost minimizing inputs are x^* $= x^t(w,p/r)$. For these inputs maximum revenue is $R^t(x^*,p) = r$, and the implied revenue maximizing outputs are $y^* = y^t(x^*,p/r) =$ $y^t(w,p/r)$. Notice that $(p/r)y^* = 1$. Notice further that the fourth expression of (2.3.10) tells us that $IC^t(w,p/r)$ can be seen as the outcome of a two stage process. In the first stage one looks for the minimum cost of producing outputs y, and in the second stage one optimizes over these costs, using the constraint $py \geq r$. The indirect cost function is linearly homogeneous in input prices w and nonincreasing in normalized output prices p/r.

(2.3.11) **Proposition**: The period t technology exhibits global CRS if and only if the indirect cost function is homogeneous of degree -1 in p/r.

Proof: Under CRS the revenue function is linearly homogeneous in the input quantity vector x. Then

$$\begin{aligned} IC^t(w,\theta p/r) &= \min_x\{wx \mid R^t(x,\theta p/r) \geq 1\} \text{ (using (2.3.10))} \\ &= \min_x\{wx \mid R^t(\theta x, p/r) \geq 1\} \text{(using linear homogeneity)} \\ &= (1/\theta)\min_{\theta x}\{w(\theta x) \mid R^t(\theta x, p/r) \geq 1\} \\ &= (1/\theta)IC^t(w,p/r). \end{aligned}$$

Reversely, one uses Färe and Primont's (1995) duality (VII) to show that then $ID_i^t(x,p/r)$ is linearly homogeneous in p/r, and duality (V) to show that $D_i^t(x,y)$ is homogeneous of degree -1 in y, which is equivalent to CRS.

Q.E.D.

How does the indirect cost function behave with respect to the variable r? If the indirect cost function is continuously differentiable, applying the Envelope Theorem to the final expression in (2.3.10) yields

$$\partial IC^t(w,p/r)/\partial r = \lambda^*, \tag{2.3.12}$$

where λ^* is the Lagrangian multiplier associated with the minimization problem. A first-order condition for an interior solution is

$$w = \lambda^* \nabla_x R^t(x^*, p), \tag{2.3.13}$$

where x^* is the solution to the minimization problem. Multiplying the lefthand and righthand side of (2.3.13) by x^*, we obtain

$$IC^t(w,p/r) = wx^* = \lambda^* x^* \nabla_x R^t(x^*, p) \tag{2.3.14}$$

$$= \lambda^*[x^* \nabla_x R^t(x^*, p)/R^t(x^*, p)]r$$
$$= \lambda^* \epsilon_R^t(x^*, p) r$$

by using respectively the other first-order condition $R^t(x^*, p) = r$ and definition (2.2.20). Combining (2.3.12) and (2.3.14), we obtain

$$\partial IC^t(w, p/r)/\partial r = IC^t(w, p/r)/\epsilon_R^t(x^*, p)r. \tag{2.3.15}$$

Rewriting this expression, we obtain

$$\partial \ln IC^t(w, p/r)/\partial \ln r = 1/\epsilon_R^t(x^*, p). \tag{2.3.16}$$

Thus the elasticity of the indirect cost function with respect to revenue r is equal to the reciprocal of the elasticity of the revenue function with respect to input in the optimum point, which is, by (2.2.20), equal to the output distance function measure of local scale elasticity $\epsilon_o^t(x^*, y^*)$. A differently styled proof of this result was given by Brown and Chachere (1986). It is now immediately clear that if the period t technology exhibits CRS then $\partial \ln IC^t(w, p/r)/\partial \ln r = 1$, that is, the indirect cost function is linearly homogeneous in r. In that case $IC^t(w, p/r) = rIC^t(w, p)$. This also follows from Proposition 2.3.11.

The linear homogeneity in x of the indirect input distance function implies that $x/ID_i^t(x, p/r)$ for all $(x, p/r)$ satisfies the constraint in the cost minimization problem (2.3.10). Thus $IC^t(w, p/r) \leq wx/ID_i^t(x, p/r)$, or

$$IC^t(w, p/r) ID_i^t(x, p/r) \leq wx \text{ for all } w, x, p/r. \tag{2.3.17}$$

This is an instance of what is called Mahler's Inequality. Against the backdrop of this inequality we will define three efficiency measures. Suppose that our firm, confronted with input prices w^t and output prices p^t, uses input quantities x^t and obtains revenue r^t. The *indirect cost efficiency* is defined by the ratio of the minimum cost of 'producing' revenue r^t to the actual cost,

$$ICE^t(w^t, x^t, p^t/r^t) \equiv IC^t(w^t, p^t/r^t)/w^t x^t. \tag{2.3.18}$$

Since apparently $x^t \in IL^t(p^t/r^t)$, $IC^t(w^t,p^t/r^t) \leq w^t x^t$ and thus $ICE^t(w^t,x^t,p^t/r^t) \leq 1$. Recall that, by (2.3.4), $x^t/ID_i^t(x^t,p^t/r^t)$ is the largest contraction of x^t that is just able to 'produce' revenue r^t. Thus a natural definition of the *indirect input technical efficiency* is

$$IITE^t(x^t,p^t/r^t) \equiv 1/ID_i^t(x^t,p^t/r^t). \tag{2.3.19}$$

Since we supposed that x^t 'produced' r^t, by (2.3.5), $IITE^t(x^t,p^t/r^t) \leq 1$. Furthermore, Mahler's Inequality implies that $ICE^t(w^t,x^t,p^t/r^t) \leq IITE^t(x^t,p^t/r^t)$. Thus, the *indirect input allocative efficiency* is defined residually by

$$IIAE^t(w^t,x^t,p^t/r^t) \equiv ICE^t(w^t,x^t,p^t/r^t)/IITE^t(x^t,p^t/r^t). \tag{2.3.20}$$

It is evident that $IIAE^t(w^t,x^t,p^t/r^t) \leq 1$. Expression (2.3.20) can be reformulated as: indirect cost efficiency can be decomposed into indirect input technical efficiency and indirect input allocative efficiency.

2.4 THE INDIRECT INPUT PRICE INDEX AND QUANTITY INDEX

Based on the indirect cost function, the *indirect input price index* is defined by

$$IP_i^t(w,w',p/r) \equiv IC^t(w,p/r)/IC^t(w',p/r), \tag{2.4.1}$$

where w and w' are two input price vectors. This index was discussed by Archibald (1975) – he called it the "fixed-revenue input price index" – Fisher (1985), (1995), and Zieschang (1979). As can be verified immediately, it has the following properties: [2]

(i) If $w \geq w''$ then $IP_i^t(w,w',p/r) \geq IP_i^t(w'',w',p/r)$, and if $w' \geq w''$ then $IP_i^t(w,w',p/r) \leq IP_i^t(w,w'',p/r)$ (monotonicity).

[2] These properties correspond to well-known axioms and tests from axiomatic price and quantity index theory (see Balk 1995b).

(ii) $IP_i^t(\lambda w, w', p/r) = \lambda\, IP_i^t(w, w', p/r)$ $(\lambda > 0)$ (linear homogeneity) and $IP_i^t(w, \lambda w', p/r) = (1/\lambda)\, IP_i^t(w, w', p/r)$ $(\lambda > 0)$ (homogeneity of degree -1).

(iii) $IP_i^t(w, w, p/r) = 1$ (identity).

(iv) $IP_i^t(\lambda w, \lambda w', p/r) = IP_i^t(w, w', p/r)$ $(\lambda > 0)$ (homogeneity of degree 0).

(v) $IP_i^t(w, w', p/r)$ does not depend on the units of measurement of the quantities (dimensional invariance). This is due to the fact that the indirect cost function is an inner product of prices and quantities.

(vi) $IP_i^t(\lambda w, w, p/r) = \lambda$ and $IP_i^t(w, \lambda w, p/r) = (1/\lambda)$ $(\lambda > 0)$ (proportionality). In particular, for $N = 1$, $IP_i^t(w, w', p/r) = w/w'$.

(vii) $IP_i^t(w, w', p/r) IP_i^t(w', w'', p/r) = IP_i^t(w, w'', p/r)$ (transitivity for fixed t and p/r).

(viii) $IP_i^t(w, w', p/r) = 1/IP_i^t(w', w, p/r)$ (time reversal).

The following Proposition shows under which circumstances the dependency of the indirect input price index on p/r vanishes.

(2.4.2) **Proposition**: $IP_i^t(w, w', p/r)$ is independent of p/r if and only if the period t technology exhibits input homotheticity.

Proof: If the technology exhibits input homotheticity then, using (2.2.12) and the fourth expression in (2.3.10),

$$
\begin{aligned}
IC^t(w, p/r) &= \min_y \{H^t(y) C^t(w, 1_M) \mid py \geq r\} \\
&= IH^t(p/r) C^t(w, 1_M),
\end{aligned}
$$

and $IP_i^t(w, w', p/r) = C^t(w, 1_M)/C^t(w', 1_M)$, independent of p/r. Reversely, assume that $IP_i^t(w, w', p/r)$ is independent of p/r. Then, following the same reasoning as in the proof of Proposition 2.2.14, we conclude that $IC^t(w,p/r) = IC^t(w,1_M)IH^t(p/r)$ for some non-increasing function $IH^t(p/r)$. Using Färe and Primont's (1995) duality (VII), we obtain

$$\begin{aligned} ID_i^t(x,p/r) &= \inf_w\{wx \mid IC^t(w,1_M)IH^t(p/r) \geq 1\} \\ &= ID_i^t(x,1_M)/IH^t(p/r). \end{aligned}$$

Using now Färe and Primont's (1995) duality (V), we obtain

$$\begin{aligned} D_i^t(x,1_M) &= \inf_{p/r}\{ID_i^t(x,p/r) \mid p1_M \geq r\} \\ &= \inf_{p/r}\{ID_i^t(x,1_M)/IH^t(p/r) \mid p1_M \geq r\} \\ &= ID_i^t(x,1_M)/\sup_{p/r}\{IH^t(p/r) \mid p1_M \geq r\} \\ &= ID_i^t(x,1_M)/IH^t(1_M). \end{aligned}$$

Combining both results, we obtain

$$ID_i^t(x,p/r) = D_i^t(x,1_M)IH^t(1_M)/IH^t(p/r),$$

which is by Proposition 2.3.9 equivalent to input homotheticity.

Q.E.D.

Combining Proposition 2.4.2 with Proposition 2.2.14 we obtain the following

(2.4.3) **Corollary**: The period t technology exhibits input homotheticity if and only if $IP_i^t(w,w',p/r) = P_i^t(w,w',y)$ for all p/r, y.

Thus under input homotheticity there is no distinction between direct and indirect price indexes.

We now consider our firm. Let the base period data be (x^0,y^0,w^0,p^0) and the comparison period data be (x^1,y^1,w^1,p^1). Thus base period revenue is $p^0y^0 \equiv r^0$ and comparison period revenue is $p^1y^1 \equiv$

r^1. We assume that in both periods the firm's objective is to attain the revenue r^t $(t = 0, 1)$ with minimal cost, given input and output prices. However, we allow for inefficiency. Thus we maintain the following behavioral assumption

$$w^t x^t = IC^t(w^t, p^t/r^t) ID_i^t(x^t, p^t/r^t)\ (t = 0, 1). \tag{2.4.4}$$

Recalling (2.3.18)-(2.3.20), we see that in both periods we assume the firm to be indirect allocatively efficient, but not necessarily indirect technically efficient with respect to inputs. Otherwise said, the input quantities x_n^t $(n = 1, ..., N)$ show the 'right' proportions but not necessarily the 'right' level.

Going from period 0 to period 1, natural indirect input price index numbers are $IP_i^0(w^1, w^0, p^0/r^0)$ and $IP_i^1(w^1, w^0, p^1/r^1)$. Both compare the input price vector w^1 to the input price vector w^0. The first index number uses what we may call the Laspeyres perspective, and the second uses the Paasche perspective. Using Mahler's Inequality (2.3.17) and assumption (2.4.4), we obtain

$$\begin{aligned} IP_i^0(w^1, w^0, p^0/r^0) &= IC^0(w^1, p^0/r^0)/IC^0(w^0, p^0/r^0) \\ &\leq \frac{[w^1 x^0 / ID_i^0(x^0, p^0/r^0)]}{[w^0 x^0 / ID_i^0(x^0, p^0/r^0)]} \\ &= w^1 x^0 / w^0 x^0 \equiv P^L(w^1, x^1, w^0, x^0). \end{aligned} \tag{2.4.5}$$

Similarly, we obtain

$$IP_i^1(w^1, w^0, p^1/r^1) \geq w^1 x^1 / w^0 x^1 \equiv P^P(w^1, x^1, w^0, x^0). \tag{2.4.6}$$

Thus, $IP_i^0(w^1, w^0, p^0/r^0)$ is bounded from above by the ordinary Laspeyres input price index number $P^L(w^1, x^1, w^0, x^0)$, and $IP_i^1(w^1, w^0, p^1/r^1)$ is bounded from below by the ordinary Paasche input price index number $P^P(w^1, x^1, w^0, x^0)$. If the discrepancies are not too large we may combine both inequalities, which gives [3]

$$[IP_i^0 IP_i^1]^{1/2} \simeq [P^L(w^1, x^1, w^0, x^0) P^P(w^1, x^1, w^0, x^0)]^{1/2}$$

[3] This depends, here and in the forthcoming cases, on the following reasoning. If $(A - a)/A \simeq (b - B)/B$ then $(AB)^{1/2}$ is a reasonable first order approximation of $(ab)^{1/2}$. This can be shown by simple Taylor expansions. If $(A - a)/A = (b - B)/B$ it is even a second order approximation.

$$\equiv P^F(w^1, x^1, w^0, x^0), \tag{2.4.7}$$

where in the lefthand side the arguments have been suppressed to avoid clumsiness of notation. Thus the geometric average of IP_i^0 and IP_i^1 may be approximated by the ordinary Fisher input price index number $P^F(w^1, x^1, w^0, x^0)$.

The foregoing result did not require any specific knowledge about the base period and comparison period technologies. Let us now assume that the indirect cost functions have the translog form, that is

$$\begin{aligned}\ln IC^t(w, p/r) \equiv{}& \alpha_0^t + \sum_{n=1}^{N} \alpha_n^t \ln w_n + \sum_{m=1}^{M} \beta_m^t \ln(p_m/r) \\ &+\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} \alpha_{nn'}^t \ln w_n \ln w_{n'} \\ &+\frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} \beta_{mm'}^t \ln(p_m/r)\ln(p_{m'}/r) \\ &+\sum_{n=1}^{N}\sum_{m=1}^{M} \gamma_{nm}^t \ln w_n \ln(p_m/r) (t = 0, 1)\end{aligned} \tag{2.4.8}$$

with the following restrictions due to linear homogeneity in input prices:

$$\sum_{n=1}^{N} \alpha_n^t = 1, \sum_{n'=1}^{N} \alpha_{nn'}^t = 0 \; (n = 1, ..., N), \tag{2.4.9}$$

$$\sum_{n=1}^{N} \alpha_{nn'}^t = 0 \; (n' = 1, ..., N), \sum_{n=1}^{N} \gamma_{nm}^t = 0 \; (m = 1, ..., M) \; (t = 0, 1).$$

If, in addition to (2.4.9), we assume that the second-order coefficients of the input prices are time-invariant, that is

$$\alpha_{nn'}^0 = \alpha_{nn'}^1 \; (n, n' = 1, ..., N), \tag{2.4.10}$$

then we obtain by using the Translog Identity (see the Appendix of Caves, Christensen and Diewert 1982)

$$\begin{aligned}
\ln[IP_i^0 IP_i^1]^{1/2} &= \frac{1}{2}[\ln IP_i^0(w^1, w^0, p^0/r^0) + \ln IP_i^1(w^1, w^0, p^1/r^1)] \\
&= \frac{1}{2}[\ln IC^0(w^1, p^0/r^0) - \ln IC^0(w^0, p^0/r^0) \\
&\quad + \ln IC^1(w^1, p^1/r^1) - \ln IC^1(w^0, p^1/r^1)] \\
&= \frac{1}{2}[\nabla_{\ln w} \ln IC^0(w^0, p^0/r^0) \\
&\quad + \nabla_{\ln w} \ln IC^1(w^1, p^1/r^1)][\ln w^1 - \ln w^0], \quad (2.4.11)
\end{aligned}$$

where $\ln w \equiv (\ln w_1, ..., \ln w_N)$. Shephard's Lemma applied to (2.4.4) yields

$$\partial IC^t(w^t, p^t/r^t)/\partial w_n = x_n^t / ID_i^t(x^t, p^t/r^t) \ (n = 1, ..., N; t = 0, 1), \quad (2.4.12)$$

which implies, using (2.4.4) again,

$$\begin{aligned}
\partial \ln IC^t(w^t, p^t/r^t)/\partial \ln w_n &= \frac{w_n^t}{IC^t(w^t, p^t/r^t)} \frac{\partial IC^t(w^t, p^t/r^t)}{\partial w_n} \\
&= w_n^t x_n^t / w^t x^t \\
&\equiv s_n^t \ (n = 1, ..., N; t = 0, 1). \quad (2.4.13)
\end{aligned}$$

Substituting (2.4.13) into (2.4.11) we obtain the following result:

(2.4.14) **Proposition**: If the technologies of base period and comparison period are characterized by translog indirect cost functions (2.4.8)-(2.4.9) with identical second-order coefficients of the input prices, that is (2.4.10) holds, and it is assumed that (2.4.4) holds, then $[IP_i^0 IP_i^1]^{1/2} = P^T(w^1, x^1, w^0, x^0)$,

where $P^T(w^1, x^1, w^0, x^0)$ is the Törnqvist input price index number which is defined by $P^T(w^1, x^1, w^0, x^0) \equiv \exp[\frac{1}{2}\sum_{n=1}^N (s_n^0 + s_n^1) \ln(w_n^1/w_n^0)]$. Notice that each translog function (2.4.8) is a member of the class of flexible functional forms, but that assumption (2.4.10) restricts the flexibility of either $IC^0(w,p/r)$ or $IC^1(w,p/r)$.

Based on the indirect input distance function the *indirect input quantity index* is defined by

$$IQ_i^t(x,x',p/r) \equiv ID_i^t(x,p/r)/ID_i^t(x',p/r), \tag{2.4.15}$$

where x and x' are two input quantity vectors. Recall that $ID_i^t(x,p/r)$ is a radial measure of the distance of x to the boundary of $IL^t(p/r)$. Thus $IQ_i^t(x,x',p/r)$ is the ratio of the distance of x and the distance of x' to the same boundary. This index was discussed by Zieschang (1979). It has the following properties:

(i) If $x \geq x''$ then $IQ_i^t(x,x',p/r) \geq IQ_i^t(x'',x',p/r)$, and if $x' \geq x''$ then $IQ_i^t(x,x',p/r) \leq IQ_i^t(x,x'',p/r)$ (monotonicity).

(ii) $IQ_i^t(\lambda x,x',p/r) = \lambda IQ_i^t(x,x',p/r)$ $(\lambda > 0)$ (linear homogeneity), and $IQ_i^t(x,\lambda x',p/r) = (1/\lambda)IQ_i^t(x,x',p/r)$ $(\lambda > 0)$ (homogeneity of degree -1).

(iii) $IQ_i^t(x,x,p/r) = 1$ (identity).

(iv) $IQ_i^t(\lambda x,\lambda x',p/r) = IQ_i^t(x,x',p/r)$ $(\lambda > 0)$ (homogeneity of degree 0).

(v) $IQ_i^t(x,x',p/r)$ does not depend on the units of measurement of the quantities (dimensional invariance). This is due to the fact that by Färe and Primont's (1995) duality (II) of the input distance function can also be written as an inner product of prices and quantities.

(vi) $IQ_i^t(\lambda x,x,p/r) = \lambda$ and $IQ_i^t(x,\lambda x,p/r) = (1/\lambda)$ $(\lambda > 0)$ (proportionality). In particular, for $N = 1$, $IQ_i^t(x,x',p/r) = x/x'$.

(vii) $IQ_i^t(x,x',p/r)IQ_i^t(x',x'',p/r) = IQ_i^t(x,x'',p/r)$ (transitivity for fixed t and p/r).

(viii) $IQ_i^t(x,x',p/r) = 1/IQ_i^t(x',x,p/r)$ (time reversal).

Parallel to Proposition 2.4.2 we have the following

(2.4.16) **Proposition**: $IQ_i^t(x, x', p/r)$ is independent of p/r if and only if the period t technology exhibits input homotheticity.

Proof: If the technology exhibits input homotheticity then, by Proposition 2.3.9, $ID_i^t(x, p/r) = D_i^t(x, 1_M)/IH^t(p/r)$ for some non-increasing function $IH^t(p/r)$. Then $IQ_i^t(x, x', p/r) = D_i^t(x, 1_M)/D_i^t(x', 1_M)$, independent of p/r. Reversely, if $IQ_i^t(x, x', p/r)$ is independent of p/r then, by following the same reasoning as in the proof of Proposition 2.2.16, it must be the case that $ID_i^t(x, p/r) = ID_i^t(x, 1_M)/IH^t(p/r)$ for some function $IH^t(p/r)$. In the final part of the proof of Proposition 2.4.2 this is shown to entail input homotheticity.

Q.E.D.

Combining this result with Proposition 2.2.16, we obtain the following

(2.4.17) **Corollary**: The period t technology exhibits input homotheticity if and only if $IQ_i^t(x, x', p/r) = Q_i^t(x, x', y)$ for all $p/r, y$.

Thus under input homotheticity indirect and direct quantity indexes coincide. If the period t technology exhibits global CRS then, by (2.3.8),

$$IQ_i^t(x, x', p/r) = R^t(x, p)/R^t(x', p), \qquad (2.4.18)$$

that is, the indirect input quantity index is equal to the input quantity index based on the (direct) revenue function. Notice that in the case of global CRS, the revenue function is linearly homogeneous in inputs.

For our firm, natural indirect input quantity index numbers are $IQ_i^0(x^1, x^0, p^0/r^0)$ and $IQ_i^1(x^1, x^0, p^1/r^1)$. Using our basic assumption (2.4.4) and Mahler's Inequality (2.3.17), we obtain the following bounds:

$$IQ_i^0(x^1, x^0, p^0/r^0) \leq w^0x^1/w^0x^0 \equiv Q^L(w^1, x^1, w^0, x^0) \quad (2.4.19)$$

and

$$IQ_i^1(x^1, x^0, p^1/r^1) \geq w^1x^1/w^1x^0 \equiv Q^P(w^1, x^1, w^0, x^0), \quad (2.4.20)$$

where $Q^L(w^1, x^1, w^0, x^0)$ and $Q^P(w^1, x^1, w^0, x^0)$ are the ordinary Laspeyres and Paasche input quantity index numbers respectively. Combining both inequalities, we obtain the following approximation

$$\begin{aligned} [IQ_i^0 IQ_i^1]^{1/2} &\simeq [Q^L(w^1, x^1, w^0, x^0)Q^P(w^1, x^1, w^0, x^0)]^{1/2} \\ &\equiv Q^F(w^1, x^1, w^0, x^0). \qquad (2.4.21) \end{aligned}$$

Thus the geometric average of IQ_i^0 and IQ_i^1 may be approximated by the ordinary Fisher input quantity index number $Q^F(w^1, x^1, w^0, x^0)$.

Let us now assume that the technologies of both periods may be characterized by translog indirect input distance functions, that is

$$\begin{aligned} \ln ID_i^t(x, p/r) \equiv\ & \alpha_0^t + \sum_{n=1}^{N} \alpha_n^t \ln x_n + \sum_{m=1}^{M} \beta_m^t \ln(p_m/r) \quad (2.4.22) \\ & + \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \alpha_{nn'}^t \ln x_n \ln x_{n'} \\ & + \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \beta_{mm'}^t \ln(p_m/r) \ln(p_{m'}/r) \\ & + \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{nm}^t \ln x_n \ln(p_m/r) \ (t = 0, 1) \\ & (x \in \Re_{++}^N) \end{aligned}$$

with the following restrictions due to linear homogeneity in input quantities:

$$\sum_{n=1}^{N} \alpha_n^t = 1, \sum_{n'=1}^{N} \alpha_{nn'}^t = 0 \ (n = 1, ..., N), \quad (2.4.23)$$

$$\sum_{n=1}^{N} \alpha_{nn'}^{t} = 0\ (n' = 1, ..., N), \sum_{n=1}^{N} \gamma_{nm}^{t} = 0\ (m = 1, ..., M)\ (t = 0, 1).$$

If, in addition to (2.4.23), we assume that the second-order coefficients of the input quantities are time-invariant, that is

$$\alpha_{nn'}^{0} = \alpha_{nn'}^{1}\ (n, n' = 1, ..., N), \tag{2.4.24}$$

then we obtain by using the Translog Identity

$$\begin{aligned}
\ln[IQ_i^0 IQ_i^1]^{1/2} &= \frac{1}{2}[\ln IQ_i^0(x^1, x^0, p^0/r^0) + \ln IQ_i^1(x^1, x^0, p^1/r^1)] \\
&= \frac{1}{2}[\ln ID_i^0(x^1, p^0/r^0) - \ln ID_i^0(x^0, p^0/r^0) \\
&\quad + \ln ID_i^1(x^1, p^1/r^1) - \ln ID_i^1(x^0, p^1/r^1)] \\
&= \frac{1}{2}[\nabla_{\ln x} \ln ID_i^0(x^0, p^0/r^0) \\
&\quad + \nabla_{\ln x} \ln ID_i^1(x^1, p^1/r^1)][\ln x^1 - \ln x^0]. \qquad (2.4.25)
\end{aligned}$$

For evaluating the derivatives, recall that $IC^t(w, p/r) = \min_x\{wx/ID_i^t(x, p/r)\}$ which is equivalent to

$$\ln IC^t(w, p/r) = \min_x\{\ln(wx) - \ln ID_i^t(x, p/r)\}. \tag{2.4.26}$$

The first-order conditions are

$$w_n/wx^* = \partial \ln ID_i^t(x^*, p/r)/\partial x_n\ (n = 1, ..., N) \tag{2.4.27}$$

where x^* is the solution to the minimization problem. Now our assumption (2.4.4) implies that the solution to the cost minimization problem inherent in $IC^t(w^t, p^t/r^t)$ is $x^t/ID_i^t(x^t, p^t/r^t)$ $(t = 0, 1)$. Substituting this into (2.4.27) and using the linear homogeneity of $ID_i^t(x, p/r)$, we obtain

$$\begin{aligned}
\partial \ln ID_i^t(x^t, p^t/r^t)/\partial \ln x_n &= w_n^t x_n^t / w^t x^t \qquad (2.4.28) \\
&\equiv s_n^t (n = 1, ..., N; t = 0, 1).
\end{aligned}$$

Substituting (2.4.28) into (2.4.25) we obtain the following result:

(2.4.29) **Proposition**: If the technologies of base period and comparison period are characterized by translog indirect input distance functions (2.4.22)-(2.4.23) with identical second-order coefficients of the input quantities, that is (2.4.24) holds, and it is assumed that (2.4.4) holds, then $[IQ_i^0 IQ_i^1]^{1/2} = Q^T(w^1, x^1, w^0, x^0)$,

where $Q^T(w^1, x^1, w^0, x^0)$ is the Törnqvist input quantity index number defined by $Q^T(w^1, x^1, w^0, x^0) \equiv \exp[\frac{1}{2}\sum_{n=1}^N (s_n^0 + s_n^1)\ln(x_n^1/x_n^0)]$. Notice that the input quantities are required to be strictly positive, and that assumption (2.4.24) restricts the flexibility of either $ID_i^0(x, p/r)$ or $ID_i^1(x, p/r)$.

On the relation between indirect input price and quantity indexes we can say the following. For our firm we can derive, by using again assumption (2.4.4) and Mahler's Inequality (2.3.17), that

$$IP_i^0 IQ_i^0 \leq w^1x^1/w^0x^0 \leq IP_i^1 IQ_i^1, \tag{2.4.30}$$

which suggests that

$$[IP_i^0 IP_i^1]^{1/2}[IQ_i^0 IQ_i^1]^{1/2} \simeq w^1x^1/w^0x^0. \tag{2.4.31}$$

Thus the cost ratio w^1x^1/w^0x^0 can be decomposed approximately into the product of a geometric average of two indirect input price index numbers and a geometric average of two indirect input quantity index numbers. In other words, $[IP_i^0 IP_i^1]^{1/2}$ and $[IQ_i^0 IQ_i^1]^{1/2}$ satisfy approximately the product test, familiar from axiomatic index theory (see Balk 1995b). Recalling (2.4.7) and (2.4.21), we notice that for the Fisher index numbers, which approximate $[IP_i^0 IP_i^1]^{1/2}$ and $[IQ_i^0 IQ_i^1]^{1/2}$, the product test is satisfied exactly.

Finally, it is interesting to compare the indirect price and quantity index numbers to appropriate direct price and quantity index numbers. Since $p^ty^t = r^t$ $(t = 0, 1)$ we have, by using the fourth expression of (2.3.10) and (2.3.6) respectively,

$$IC^t(w, p^t/r^t) \leq C^t(w, y^t) \text{ for all } w \ (t = 0, 1) \tag{2.4.32}$$

and

$$ID_i^t(x, p^t/r^t) \geq D_i^t(x, y^t) \text{ for all } x \ (t = 0, 1). \tag{2.4.33}$$

Inequality (2.4.32) means that the minimum cost of producing y^t, and thereby attaining revenue r^t, is not necessarily the minimum cost of 'producing' revenue r^t. A similar interpretation holds for (2.4.33). In particular these inequalities hold for $w = w^t$ and $x = x^t$ respectively ($t = 0, 1$). Now suppose that $IC^t(w^t, p^t/r^t) = C^t(w^t, y^t)$ and $ID_i^t(x^t, p^t/r^t) = D_i^t(x^t, y^t)$ for $t = 0{,}1$. Then one obtains the following inequalities:

$$IP_i^0(w^1, w^0, p^0/r^0) \leq P_i^0(w^1, w^0, y^0) \tag{2.4.34}$$

$$IP_i^1(w^1, w^0, p^1/r^1) \geq P_i^1(w^1, w^0, y^1). \tag{2.4.35}$$

Thus the Laspeyres-perspective indirect input price index number is smaller than or equal to the Laspeyres-perspective direct input price index number, and the reverse relation holds for the Paasche-perspective input price index numbers. However, the combination of both inequalities suggests that

$$[IP_i^0 IP_i^1]^{1/2} \simeq [P_i^0 P_i^1]^{1/2}. \tag{2.4.36}$$

Similarly, we obtain for the input quantity index numbers

$$IQ_i^0(x^1, x^0, p^0/r^0) \geq Q_i^0(x^1, x^0, y^0) \tag{2.4.37}$$

$$IQ_i^1(x^1, x^0, p^1/r^1) \leq Q_i^1(x^1, x^0, y^1), \tag{2.4.38}$$

and

$$[IQ_i^0 IQ_i^1]^{1/2} \simeq [Q_i^0 Q_i^1]^{1/2}. \tag{2.4.39}$$

Thus the Laspeyres-perspective indirect input quantity index number is larger than or equal to the Laspeyres-perspective direct input quantity index number, and the reverse holds for the Paasche-perspective input quantity index numbers. But the geometric average of Laspeyres-perspective and Paasche-perspective indirect input quantity index numbers approximates the geometric average of their direct counterparts. The bottom line is that, using only single firm data, it is hard to distinguish empirically between geometric averages of direct and indirect price and quantity index numbers.

2.5 THE INDIRECT INPUT BASED PRODUCTIVITY INDEXES

In measuring productivity change we will concentrate on two independently operating factors, viz. technical (technological) change and change of the firm's technical efficiency. Based upon (2.3.19) we measure *indirect input technical efficiency change* between periods 0 and 1 by

$$IEC_i(x^1, p^1/r^1, x^0, p^0/r^0) \equiv ID_i^0(x^0, p^0/r^0)/ID_i^1(x^1, p^1/r^1). \tag{2.5.1}$$

The *primal indirect input-based technical change index* is defined by

$$ITC(IL^t, IL^{t'}; x, p/r) \equiv ID_i^t(x, p/r)/ID_i^{t'}(x, p/r), \tag{2.5.2}$$

and the *dual indirect input-based technical change index* is defined by

$$ITC(IL^t, IL^{t'}; w, p/r) \equiv IC^{t'}(w, p/r)/IC^t(w, p/r). \tag{2.5.3}$$

The primal index compares the distances between x and the boundaries of $IL^t(p/r)$ and $IL^{t'}(p/r)$. If, going from t' to t, the distance increases (decreases) then the boundary of the indirect input set corresponding to p/r moves toward (away from) the origin, which means that there is technical progress (regress). The dual index compares the minimum costs of 'producing' revenue r when prices are (w, p). If, going from t' to t, the minimum cost decreases (increases) then there is technical progress (regress). Both measures compare the technologies of t' and t through the use of auxiliary variables $(x, p/r)$ and $(w, p/r)$ respectively. For our firm, natural choices for these variables would be their base period and comparison period values. Since we have no preference for either of them, we opt for the geometric average of the corresponding index numbers.

The *(primal) Malmquist indirect (or revenue constrained) input based productivity index number* for period 1 relative to period 0 combines the technical efficiency change index number (2.5.1) with combines the technical efficiency change index number (2.5.1) with the geometric average of two primal technical change index numbers (2.5.2), and is defined by

$$\begin{aligned} IM_i(x^1,p^1/r^1,x^0,p^0/r^0) &\equiv IEC_i(x^1,p^1/r^1,x^0,p^0/r^0)\times \\ &\quad [ITC(IL^1,IL^0;x^0,p^0/r^0)\times \\ &\quad ITC(IL^1,IL^0;x^1,p^1/r^1)]^{1/2} \\ &= \left[\frac{ID_i^0(x^0,p^0/r^0)}{ID_i^0(x^1,p^1/r^1)}\frac{ID_i^1(x^0,p^0/r^0)}{ID_i^1(x^1,p^1/r^1)}\right]^{1/2}. \end{aligned} \tag{2.5.4}$$

Notice that a value larger than 1 denotes improvement and a value less than 1 deterioration. $IM_i(x^1,p^1/r^1,x^0,p^0/r^0)$ was defined as the reciprocal of the expression behind the equality sign by Färe and Grosskopf (1994, 136).

For a first type of nonparametric approximation we assume that the base period and comparison period technology exhibit global CRS. Then, by (2.3.8),

$$IM_i(x^1,p^1/r^1,x^0,p^0/r^0) = \left[\frac{R^0(x^0,p^0)}{R^0(x^1,p^1)}\frac{R^1(x^0,p^0)}{R^1(x^1,p^1)}\right]^{1/2}/(r^0/r^1). \tag{2.5.5}$$

Parallelling Balk (1995a), we assume further that the firm solves the profit maximization problem

$$\max_x\{R^t(x,p^t)-w^tx\}\ (t=0,1). \tag{2.5.6}$$

Then, since global CRS holds,

$$R^t(x^t,p^t)=w^tx^t\ (t=0,1), \tag{2.5.7}$$

but also

$$\max_x\{R^0(x,p^1)-w^0x\}=0. \tag{2.5.8}$$

Thus, in particular, $R^0(x^1,p^1)-w^0x^1\leq 0$, or

$$R^0(x^1,p^1)\leq w^0x^1. \tag{2.5.9}$$

Similarly we obtain

$$R^1(x^0,p^0) \leq w^1x^0. \tag{2.5.10}$$

Thus, combining (2.5.9) and (2.5.7) for $t = 0$, we obtain

$$R^0(x^1,p^1)/R^0(x^0,p^0) \leq w^0x^1/w^0x^0, \tag{2.5.11}$$

and, combining (2.5.10) and (2.5.17) for $t = 1$, we obtain

$$R^1(x^1,p^1)/R^1(x^0,p^0) \geq w^1x^1/w^1x^0. \tag{2.5.12}$$

Combining (2.5.11) and (2.5.12) in the usual way, and substituting the result into (2.5.5), we obtain finally

$$IM_i(x^1,p^1/r^1,x^0,p^0/r^0) \simeq (r^1/r^0)/Q^F(w^1,x^1,w^0,x^0). \tag{2.5.13}$$

Thus if the technologies of the base and comparison period exhibit global CRS and the firm acts profit maximizing in both periods, then the Malmquist indirect input based productivity index number is approximately equal to the actual revenue index number divided by the ordinary Fisher input quantity index number.

We now assume that both period's indirect input distance functions have the translog form, that is, (2.4.22) and (2.4.23) hold. In addition we assume that all second-order coefficients are time-invariant, that is

$$\begin{gathered} \alpha^0_{nn'} = \alpha^1_{nn'}, \beta^0_{mm'} = \beta^1_{mm'}, \\ \gamma^0_{nm} = \gamma^1_{nm} \ (n,n' = 1,...,N; m,m' = 1,...,M). \end{gathered} \tag{2.5.14}$$

Using the Translog Identity we then obtain

$$\begin{aligned} &\ln IM_i(x^1,p^1/r^1,x^0,p^0/r^0) \\ &= \frac{1}{2}[\ln ID^0_i(x^0,p^0/r^0) - \ln ID^0_i(x^1,p^1/r^1) + \ln ID^1_i(x^0,p^0/r^0) \\ &\quad - \ln ID^1_i(x^1,p^1/r^1)] \\ &= \frac{1}{2}[\nabla_{\ln x} \ln ID^0_i(x^0,p^0/r^0) + \nabla_{\ln x} \ln ID^1_i(x^1,p^1/r^1)][\ln x^0 - \ln x^1] \\ &\quad + \frac{1}{2}[\nabla_{\ln(p/r)} \ln ID^0_i(x^0,p^0/r^0) + \nabla_{\ln(p/r)} \ln ID^1_i(x^1,p^1/r^1)] \\ &\quad [\ln(p^0/r^0) - \ln(p^1/r^1)]. \end{aligned} \tag{2.5.15}$$

Recall (2.4.28) for $\nabla_{\ln x} \ln ID_i^t(x^t, p^t/r^t)$ $(t = 0, 1)$. The remaining derivatives can be found as follows. As we have seen, the indirect cost function can be expressed in several ways. The final line of (2.3.10) can be written as $IC^t(w, p/r) = \min_x\{wx \mid R^t(x, p/r) \geq 1\}$. Applying the Envelope Theorem to this expression yields

$$\nabla_{p/r} IC^t(w, p/r) = -\lambda^* \nabla_{p/r} R^t(x^*, p/r), \tag{2.5.16}$$

where x^* is the solution to the minimization problem and λ^* the associated Lagrangian multiplier. A first-order condition for an interior solution is

$$w = \lambda^* \nabla_x R^t(x^*, p/r). \tag{2.5.17}$$

Multiplying the lefthand and righthand side of this equality by x^*, using definition (2.2.20) and the other first-order condition $R^t(x^*, p/r) = 1$, we obtain

$$IC^t(w, p/r) = wx^* = \lambda^* \epsilon_R^t(x^*, p). \tag{2.5.18}$$

Combining (2.5.16) and (2.5.18) yields

$$\nabla_{p/r} IC^t(w, p/r) = -\nabla_{p/r} R^t(x^*, p/r) IC^t(w, p/r)/\epsilon_R^t(x^*, p). \tag{2.5.19}$$

We now apply the Envelope Theorem to the second line of (2.3.10). This yields

$$\nabla_{p/r} IC^t(w, p/r) = -\lambda^{**} \nabla_{p/r} ID_i^t(x^*, p/r), \tag{2.5.20}$$

where x^* is the solution to the minimization problem and λ^{**} the associated Lagrangian multiplier. A first-order condition for an interior solution is

$$w = \lambda^{**} \nabla_x ID_i^t(x^*, p/r). \tag{2.5.21}$$

Multiplying the lefthand and righthand side of this equality by x^*, using the linear homogeneity in x of the indirect input distance function and the other first-order condition $ID_i^t(x^*, p/r) = 1$, we obtain

$$IC^t(w, p/r) = wx^* = \lambda^{**}. \tag{2.5.22}$$

Combining (2.5.20) and (2.5.22) yields

$$\nabla_{p/r} IC^t(w, p/r) = -\nabla_{p/r} ID_i^t(x^*, p/r) IC^t(w, p/r). \tag{2.5.23}$$

Combining finally (2.5.19) and (2.5.23), we obtain

$$\nabla_{p/r} ID_i^t(x^*, p/r) = \nabla_{p/r} R^t(x^*, p/r)/\epsilon_R^t(x^*, p), \tag{2.5.24}$$

where x^* is the solution to the minimization problem defining $IC^t(w, p/r)$. Notice that, by Shephard's Lemma applied to the revenue function,

$$\nabla_{p/r} R^t(x^*, p/r) = y^t(x^*, p/r), \tag{2.5.25}$$

where $y^t(x^*, p/r)$ is the revenue maximizing output vector for which, since $R^t(x^*, p/r) = 1$, $(p/r)y^t(x^*, p/r) = 1$ holds. Now our basic assumption (2.4.4) implies that the solution to the cost minimization problem inherent in $IC^t(w^t, p^t/r^t)$ is given by the vector $x^t/ID_i^t(x^t, p^t/r^t)$ $(t = 0, 1)$. Thus $(p^t/r^t)y^t(x^t/ID_i^t(x^t, p^t/r^t), p^t/r^t)$ $= 1$. By definition we have that $(p^t/r^t)y^t = 1$. In addition to our basic assumption we assume that

$$y^t(x^t/ID_i^t(x^t, p^t/r^t), p^t/r^t) = y^t \ (t = 0, 1), \tag{2.5.26}$$

that is, the actual outputs are optimal for the revenue constrained cost minimization. Then (2.5.24) transforms into

$$\nabla_{p/r} ID_i^t(x^t, p^t/r^t)/ID_i^t(x^t, p^t/r^t) = y^t/\epsilon_R^{t*} \tag{2.5.27}$$

where $\epsilon_R^{t*} \equiv \epsilon_R^t(x^t/ID_i^t(x^t, p^t/r^t), p^t)$ $(t = 0, 1)$, or

$$\begin{aligned} \partial \ln ID_i^t(x^t, p^t/r^t)/\partial \ln(p_m/r) &= p_m^t y_m^t / r^t \epsilon_R^{t*} \\ &= p_m^t y_m^t / p^t y^t \epsilon_R^{t*} \\ &\equiv u_m^t/\epsilon_R^{t*} \ (m = 1, ..., M; t = 0, 1). \end{aligned} \tag{2.5.28}$$

By substituting (2.4.28) and (2.5.28) into (2.5.15) we obtain

$$\begin{aligned} &\ln IM_i(x^1, p^1/r^1, x^0, p^0/r^0) \\ &= -\frac{1}{2} \sum_{n=1}^{N} (s_n^0 + s_n^1) \ln(x_n^1/x_n^0) - \frac{1}{2} \sum_{m=1}^{M} (u_m^0/\epsilon_R^{0*} + u_m^1/\epsilon_R^{1*}) \\ &\quad (\ln(p_m^1/p_m^0) - \ln(r^1/r^0)). \end{aligned} \tag{2.5.29}$$

Summarizing the foregoing, we have obtained the following result.

(2.5.30) **Proposition**: If the technologies of base period and comparison period are characterized by translog indirect input distance functions (2.4.22)-(2.4.23) with identical second-order coefficients, that is (2.5.14) holds, and it is assumed that (2.4.4) and (2.5.26) hold, then

$$IM_i(x^1, p^1/r^1, x^0, p^0/r^0) =$$

$$\frac{(r^1/r^0)^{(1/\epsilon_R^{0*}+1/\epsilon_R^{1*})}/P^{T*}(p^1, y^1, p^0, y^0)}{Q^T(w^1, x^1, w^0, x^0)},$$

where the modified Törnqvist output price index number is now defined by

$$P^{T*}(p^1, y^1, p^0, y^0) \equiv \exp[\frac{1}{2}\sum_{m=1}^{M}(u_m^0/\epsilon_R^{0*} + u_m^1/\epsilon_R^{1*})\ln(p_m^1/p_m^0)]. \tag{2.5.31}$$

Recall that, by (2.3.16), $1/\epsilon_R^{t*}$ is the elasticity of the indirect cost function with respect to revenue r, and ϵ_R^{t*} is a measure of local scale elasticity. Thus under local/global CRS, $\epsilon_R^{t*} = 1$ $(t = 0, 1)$ and the expression in Proposition 2.5.30 boils down to

$$IM_i(x^1, p^1/r^1, x^0, p^0/r^0) = \frac{(r^1/r^0)/P^T(p^1, y^1, p^0, y^0)}{Q^T(w^1, x^1, w^0, x^0)}, \tag{2.5.32}$$

that is a deflated revenue index number divided by an input quantity index number. Notice that assumption (2.5.14) restricts the flexibility of either $ID_i^0(x, p/r)$ or $ID_i^1(x, p/r)$.

We finally turn to the *dual indirect (or revenue constrained) input based productivity index number* for period 1 relative to period 0, which is, analogous to (2.5.4) but now using dual technical change index numbers, defined by

$$IM_i(w^1, x^1, p^1/r^1, w^0, x^0, p^0/r^0) \tag{2.5.33}$$

$$
\begin{aligned}
&\equiv IEC_i(x^1, p^1/r^1, x^0, p^0/r^0) \times \\
&\quad [ITC(IL^1, IL^0; w^0, p^0/r^0) ITC(IL^1, IL^0; w^1, p^1/r^1)]^{1/2} \\
&= \frac{ID_i^0(x^0, p^0/r^0)}{ID_i^1(x^1, p^1/r^1)} \left[\frac{IC^0(w^0, p^0/r^0)}{IC^1(w^0, p^0/r^0)} \frac{IC^0(w^1, p^1/r^1)}{IC^1(w^1, p^1/r^1)} \right]^{1/2} \\
&= \frac{w^0 x^0}{w^1 x^1} \left[\frac{IC^1(w^1, p^1/r^1)}{IC^1(w^0, p^0/r^0)} \frac{IC^0(w^1, p^1/r^1)}{IC^0(w^0, p^0/r^0)} \right]^{1/2},
\end{aligned}
$$

where the final expression was obtained by using assumption (2.4.4). We assume that the indirect cost functions have the translog form (2.4.8)-(2.4.9) with all second-order coefficients being time-invariant, that is

$$
\begin{aligned}
&\alpha^0_{nn'} = \alpha^1_{nn'}, \beta^0_{mm'} = \beta^1_{mm'}, \\
&\quad \gamma^0_{nm} = \gamma^1_{nm} \; (n, n' = 1, ..., N; m, m' = 1, ..., M).
\end{aligned} \tag{2.5.34}
$$

Using the Translog Identity we then obtain

$$
\begin{aligned}
&\ln \left[\frac{IC^1(w^1, p^1/r^1)}{IC^1(w^0, p^0/r^0)} \frac{IC^0(w^1, p^1/r^1)}{IC^0(w^0, p^0/r^0)} \right]^{1/2} \\
&= \frac{1}{2}[\nabla_{\ln w} \ln IC^0(w^0, p^0/r^0) + \nabla_{\ln w} \ln IC^1(w^1, p^1/r^1)] \\
&\quad [\ln w^1 - \ln w^0] \\
&\quad + \frac{1}{2}[\nabla_{\ln(p/r)} \ln IC^0(w^0, p^0/r^0) + \nabla_{\ln(p/r)} \ln IC^1(w^1, p^1/r^1)] \\
&\quad [\ln(p^1/r^1) - \ln(p^0/r^0)].
\end{aligned} \tag{2.5.35}
$$

Recall from (2.4.13) that $\nabla_{\ln w} \ln IC^t(w^t, p^t/r^t) = s^t$ $(t = 0, 1)$. Combining (2.5.23), (2.5.24) and (2.5.25) we obtain

$$
\nabla_{p/r} IC^t(w, p/r)/IC^t(w, p/r) = -y^t(x^*, p/r)/\epsilon^t_R(x^*, p), \tag{2.5.36}
$$

where x^* is the solution to the minimization problem defining $IC^t(w, p/r)$. Using assumptions (2.4.4) and (2.5.26), (2.5.36) transforms into

$$
\partial \ln IC^t(w^t, p^t/r^t)/\partial \ln(p_m/r) = -u^t_m/\epsilon^{t*}_R \; (m = 1, ..., M; t = 0, 1). \tag{2.5.37}
$$

Substituting (2.4.13) and (2.5.37) into (2.5.35), and substituting the result into (2.5.33), we obtain

$$\begin{aligned}
&\ln IM_i(w^1, x^1, p^1/r^1, w^0, x^0, p^0/r^0) \qquad (2.5.38)\\
&= \ln(w^0x^0/w^1x^1) + \frac{1}{2}\sum_{n=1}^{N}(s_n^0 + s_n^1)\ln(w_n^1/w_n^0)\\
&-\frac{1}{2}\sum_{m=1}^{M}(u_m^0/\epsilon_R^{0*} + u_m^1/\epsilon_R^{1*})(\ln(p_m^1/p_m^0) - \ln(r^1/r^0)).
\end{aligned}$$

We summarize the foregoing:

(2.5.39) **Proposition**: If the technologies of base period and comparison period are characterized by translog indirect cost functions (2.4.8)-(2.4.9) with identical second-order coefficients, that is (2.5.34) holds, and it is assumed that (2.4.4) and (2.5.26) hold, then

$$\begin{aligned}
&IM_i(w^1, x^1, p^1/r^1, w^0, x^0, p^0/r^0)\\
&= \frac{(r^1/r^0)^{(1/\epsilon_R^{0*}+1/\epsilon_R^{1*})}/P^{T*}(p^1, y^1, p^0, y^0)}{(w^1x^1/w^0x^0)/P^T(w^1, x^1, w^0, x^0)}
\end{aligned}$$

where $P^{T*}(p^1, y^1, p^0, y^0)$ was defined by (2.5.31). Notice that assumption (2.5.34) restricts the flexibility of either $IC^0(w, p/r)$ or $IC^1(w, p/r)$. Under local/ global CRS the expression in Proposition 2.5.39 boils down to

$$IM_i(w^1, x^1, p^1/r^1, w^0, x^0, p^0/r^0) = \frac{(r^1/r^0)/P^T(p^1, y^1, p^0, y^0)}{(w^1x^1/w^0x^0)/P^T(w^1, x^1, w^0, x^0)}, \qquad (2.5.40)$$

that is a deflated revenue index number divided by a deflated cost index number, where Törnqvist price index numbers are used as deflators.

2.6 CONCLUSION

A competitive, revenue-constrained firm minimizes the input cost subject to the attainment of a target revenue. Input and output prices are considered as given. According to Fisher (1995), this model can also be used for a small productive sector or for a small country involved in international trade. In our development we allowed for (radial) technical inefficiency with respect to inputs on the part of the firm. The technology of a revenue-constrained firm can be described by an indirect input distance function or an indirect cost function. These functions are the appropriate building blocks for input price and quantity indexes. These indexes will in general depend on certain reference variables.

It appeared that geometric averages of base period and comparison period referenced price and quantity index numbers can provide a complete decomposition of the input cost ratio (cf. result (2.4.31)). It also appeared that these average index numbers can be approximated by conventional Fisher index numbers (cf. results (2.4.7) and (2.4.21)). Under additional restrictions on the underlying technologies it appeared that the average index numbers are equal to conventional Törnqvist index numbers (cf. Propositions 2.4.14 and 2.4.29). It should be remarked that the conventional Laspeyres and Paasche index numbers retain their bounding properties (cf. results (2.4.5), (2.4.6), (2.4.19) and (2.4.20)), that is, they are either over- or under-estimates.

Based on the indirect distance function and the indirect cost function we defined primal and dual productivity index numbers, which capture the effect of efficiency change and technical change. Under rather mild restrictions on the underlying technologies and with help of an additional assumption about the firm's behavior, it appeared possible to arrive at computable expressions for the productivity index numbers (cf. Propositions 2.5.30 and 2.5.39). In both cases the expressions have the form of an implicit output quantity index number divided by an (implicit) input quantity index number. However, except in the case of global CRS technologies, we need the local scale elasticities pertaining to both observation periods.

References

Archibald, R. B., 1975, *On the Theory of Industrial Price Measurement: Input Price Indexes*, Working Paper No. 48 (US Bureau of Labor Statistics, Washington DC).

Balk, B. M., 1995a, "On Approximating the Indirect Malmquist Productivity Indices by Fisher Indices," *The Journal of Productivity Analysis* 6, 195-200.

Balk, B. M., 1995b, "Axiomatic Price Index Theory: A Survey," *International Statistical Review* 63, 69-93.

Brown, L. and G. R. Chachere, 1986, "A Multiproduct Cost Elasticity," *Economics Letters* 22, 45-49.

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.

Court, L. M. and H. G. Lewis, 1942-43, "Production Cost Indices," *The Review of Economic Studies* 10, 28-42.

Färe, R. and S. Grosskopf, 1994, *Cost and Revenue Constrained Production*, Bilkent University Lecture Series Volume 4 (Springer-Verlag, New York etc.).

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

Fisher, F. M., 1985, Production-Theoretic Input Price Indices and the Measurement of Real Aggregate Input Use, Paper presented at the Fourth Karlsruhe Seminar on Measurement in Economics. Sections 1-3 published in *Measurement in Economics*, edited by W. Eichhorn (Physica-Verlag, Heidelberg, 1988).

Fisher, F. M., 1995, "The Production-Theoretic Measurement of Input Price and Quantity Indices," *Journal of Econometrics* 65, 155-174.

Konüs, A. A., 1924, "The Problem of the True Index of the Cost of Living," translated in *Econometrica* 7, 1939, 10-29.

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

Shephard, R. W., 1953, *Cost and Production Functions*, reprinted as Lecture Notes in Economics and Mathematical Systems No. 194 (Springer-Verlag, Berlin etc., 1981).

Shephard, R. W., 1974, *Indirect Production Functions*, Mathematical Systems in Economics No. 10 (Anton Hain, Meisenheim am Glan).

Zieschang, K. D., 1979, *The Theory of Price and Quantity Indexes for Multiproduct Production Systems*, Working Paper No. 99 (US Bureau of Labor Statistics, Washington DC).

3

MALMQUIST PRODUCTIVITY INDEXES: A SURVEY OF THEORY AND PRACTICE

by

Rolf Färe, Shawna Grosskopf and Pontus Roos*

3.1 INTRODUCTION

In 1982, Caves, Christensen and Diewert introduced the Malmquist productivity index. Although that paper was extremely influential and widely cited, the Malmquist productivity index itself was rarely computed (Pittman, (1983) and Nishimizu and Page (1982) are exceptions), until Färe, Grosskopf, Lindgren and Roos (1989), showed how it could be calculated using a nonparametric linear programming method. In this paper we review the theoretical development and empirical applications of (some of) the various Malmquist productivity indexes.

Other surveys of productivity which include the Malmquist indexes are Diewert (1992:a,1993), Hjalmarsson (1991) in Swedish, Roos (1993) and Sudit (1995).

*Thanks are due to Balk, B., Bjurek, H., Førsund, F., Russell, R.R., Veiderpass, A. and Weber, W. for their comments.

3.2 THEORY

In this part of the paper, we study the theoretical aspects of the Malmquist productivity indexes. The basic tools are the input and output distance functions, defined by radial scalings of inputs and outputs, respectively. These functions are introduced in Section 3.2.1 and a brief discussion of their properties is provided.

Equipped with the distance functions, Section 3.2.2 turns to the definitions of the Malmquist productivity indexes. From the definition of the input quantity index by Malmquist (1953), we find that the analogous productivity indexes are those introduced by Caves, Christensen and Diewert (1982) (hereafter, CCD). These indexes are contrasted with the Hicks-Moorsteen approach which measures productivity as the ratio of quantity indexes.

In Section 3.2.3, we discuss returns to scale and Malmquist productivity measurement. In particular, we show how the one-input, one-output version of the CCD indexes are related to ratios of average products or total factor productivity. This section is followed by one that relates the Malmquist productivity indexes to the Törnqvist and Fisher productivity indexes. Of course, these latter two can only be computed if price data are available. As we shall see in Section 3.2.7, price data are not required for the Malmquist indexes.

Another important feature of the Malmquist productivity index is that it can be exhaustively partitioned into useful component measures. In particular, the index can be decomposed into a technical change and an efficiency change component. This and additional decompositions are developed in Section 3.2.5.

In Section 3.2.6, we address the issue of the circularity test. If we let t_1, t_2 and t_3 represent three periods, then an index I is said to satisfy the circularity test if $I(t_1,t_3) = I(t_1,t_2)I(t_2,t_3)$; i.e., it is transitive. In general, Malmquist indexes do not satisfy this test, so additional conditions must be imposed. We discuss these and other attempts to find a Malmquist index that satisfies the circularity test.

We conclude part 3.2 by discussing various approaches to the computation of the Malmquist productivity index. These include parametric and nonparametric linear programming approaches.

3.2.1 Distance Functions

Malmquist defined his distance function as the radial contraction to an indifference curve, while Shephard (1953) defined it in terms of a production function. Here we define the input distance function directly on the technology as in Shephard (1970).

Let $x = (x_1, \ldots, x_N) \in \Re_+^N$ denote a vector of inputs and $y = (y_1, \ldots, y_M) \in \Re_+^M$ be an output vector. The production technology T is defined by

$$T = \{(x, y) : x \text{ can produce } y\}, \tag{3.2.1}$$

and it consists of all input-output vectors that are technically feasible. In the case of a single output, the production function is often used to represent the technology. This function is defined by

$$F(x) = \max\{y : (x, y) \in T\},^{1} \tag{3.2.2}$$

and T can be recovered from F as

$$T = \{(x, y) : F(x) \geq y, x \in \Re_+^N\}. \tag{3.2.3}$$

The input distance function is defined on the technology T as

$$D_i(y, x) = \sup \left\{ \lambda : (\frac{x}{\lambda}, y) \in T \right\}, \tag{3.2.4}$$

i.e., as the "maximal" feasible contraction of x. From its definition, it follows that the input distance function is homogeneous of degree

[1] If T is a closed set and if the output set $P(x) = \{y : (x, y) \in T\}$ is bounded for each $x \in \Re_+^N$, it can then be proved that the production function exists.

+1 in inputs, i.e.,

$$D_i(y, \lambda x) = \lambda D_i(y, x), \quad \text{for all } \lambda > 0. \tag{3.2.5}$$

Moreover, if inputs are weakly disposable,[2] then it is a complete characterization of T,[3] i.e.,

$$T = \{(x, y) : D_i(y, x) \geq 1\}. \tag{3.2.6}$$

The output distance function, due to Shephard (1970) is defined by

$$D_o(x, y) = \inf\left\{\theta : \left(x, \frac{y}{\theta}\right) \in T\right\}, \tag{3.2.7}$$

or due to (2.1.6),

$$D_o(x, y) = \inf\left\{\theta : D_i\left(\frac{y}{\theta}, x\right) \geq 1\right\}. \tag{3.2.8}$$

This function is homogeneous of degree +1 in outputs, and it is a complete characterization of T provided outputs are weakly disposable,[4] i.e.,

$$T = \{(x, y) : D_o(x, y) \leq 1\}. \tag{3.2.9}$$

Thus if inputs and outputs are weakly disposable, then

$$D_o(x, y) \leq 1 \text{ if and only if } D_i(y, x) \geq 1. \tag{3.2.9$'$}$$

If the technology T exhibits constant returns to scale, one can prove that the input and output distance functions are reciprocal. In technical terms, Constant Returns to Scale (CRS) is defined as

$$\lambda T = T, \quad \text{for all } \lambda > 0. \tag{3.2.10}$$

Under this condition, we have

$$D_o(x, y) = 1/D_i(y, x), \tag{3.2.11}$$

[2]Inputs are weakly disposable if $(x, y) \in T$ and $\lambda \geq 1$ imply $(\lambda x, y) \in T$.

[3]See Färe and Primont (1995, p. 22).

[4]Outputs are weakly disposable if $(x, y) \in T$ and $0 \leq \theta \leq 1$ imply $(x, \theta y) \in T$. The proof of (3.2.9) follows from Färe and Primont (1995).

i.e., the two distance functions are reciprocals. Moreover, the output distance function is homogeneous of degree -1 in inputs and the input distance function is homogeneous of degree -1 in outputs.

To illustrate the two distance functions, assume that one input is used to produce one output. Given observed input-output combination (x, y), the two distance functions are easily illustrated as can be seen from Figure 3.1. The output distance function scales the observed point due North (in the output direction) reaching the boundary of technology. The input distance function scales the observed point due West (in the input direction) until the boundary is attained. See the paper by Russell in this volume for a more detailed discussion on distance functions.

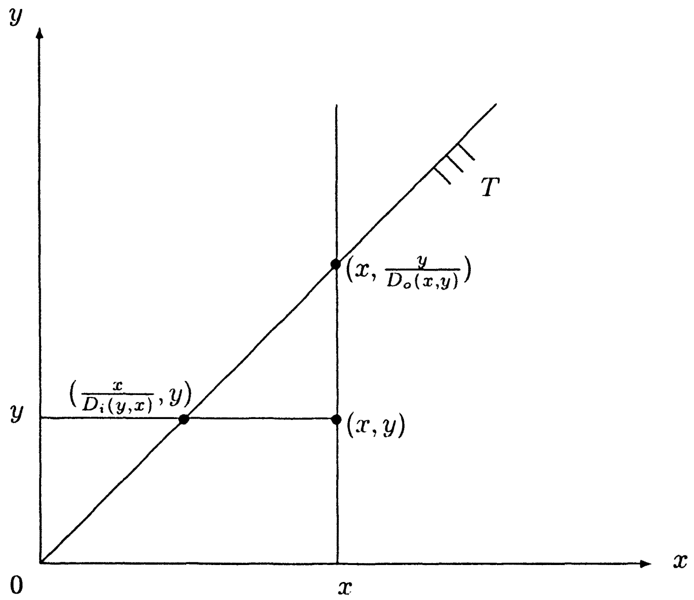

Figure 3.1 Input and Output Distance Functions

3.2.2 Malmquist Productivity Indexes: Definitions

Caves, Christensen and Diewert (1982) introduced two theoretical indexes which they named Malmquist input and output productivity indexes. These indexes follow the spirit of Sten Malmquist's (1953) quantity index. Malmquist constructed his quantity index by comparing two quantity vectors to an arbitrary indifference curve using radial scaling. Caves, Christensen and Diewert compare two input-output vectors to a reference technology using radial input and output scaling, for the input and output productivity indexes respectively.

Let t and $t+1$ denote two time periods and let $D_o^t(x^t, y^t)$ be the value of the output distance function for the technology from period t and the input-output vector from the same period. Let $D_o^t(x^{t+1}, y^{t+1})$ be the value of the distance function for the input-output vector of period $t+1$ and the technology at t. The t-period output-oriented Malmquist productivity index due to Caves, Christensen and Diewert is defined as[5]

$$M_o^t = \frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)}. \tag{3.2.12}$$

This index is illustrated in Figure 3.2.

The t-period technology is represented by T^t and the two input-output vectors by (x^t, y^t) and (x^{t+1}, y^{t+1}). (x^t, y^t) is feasible, of course, but (x^{t+1}, y^{t+1}) is outside T^t; thus $D_o^t(x^t, y^t) \leq 1$ and $D_o^t(x^{t+1}, y^{t+1}) > 1$. In terms of the distances on the y-axis, the productivity index is

$$M_o^t = \frac{Od}{Oe} / \frac{Oa}{Ob}, \tag{3.2.13}$$

and a productivity improvement would be signaled since this value

[5]The corresponding input oriented index is

$$M_i^t = D_i^t(y^{t+1}, x^{t+1}) / D_i^t(y^t, x^t).$$

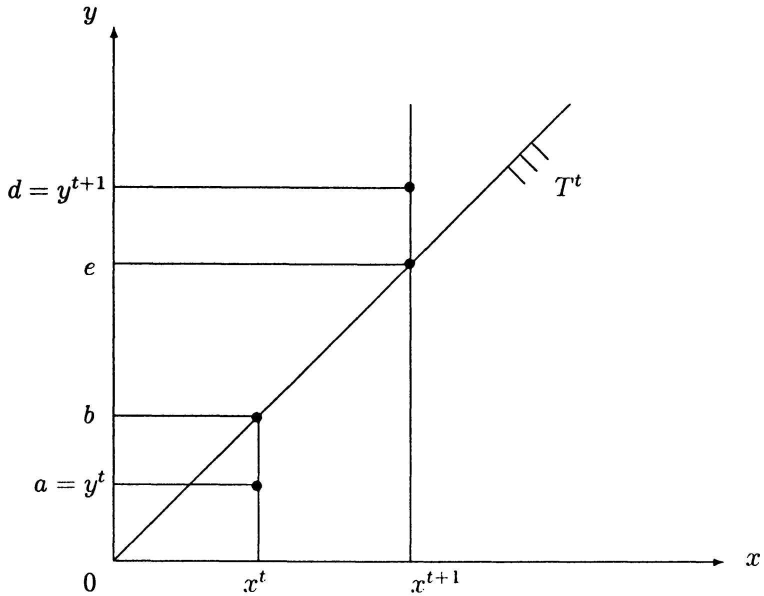

Figure 3.2 The t Period Malmquist Output Oriented Productivity Index

is larger than one.

Two time periods are involved in Definition 3.2.12 suggesting a $t+1$ productivity index namely[6]

$$M_o^{t+1} = \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^t, y^t)}. \tag{3.2.14}$$

In general, of course, M_o^t and M_o^{t+1} yield different productivity numbers since their reference technologies may differ.

It is quite easy to prove that $M_o^t = M_o^{t+1}$ if and only if the distance functions are of the form

$$D_o^\tau(x,y) = \hat{A}(\tau)\hat{D}_o(x,y), \tau = t, t+1, \ldots, \tag{3.2.15}$$

i.e., the technology is Hicks output-neutral and the output set $P^\tau(x)$

[6]Also due to Caves, Christensen and Diewert (1982). Actually, they did not just suggest t for time, they allowed for different firms as well.

equals

$$P^{\tau}(x) = (1/\hat{A}(\tau))\hat{P}(x). \tag{3.2.16}$$

Since this condition for equality between (3.2.1) and (3.2.3) is quite restrictive, we prefer not to impose it on the reference technology. In this case, one may either choose (3.2.1) or (3.2.3) to compute productivity, however the choice between the two is arbitrary. Inspired by Caves, Christensen and Diewert[7], Färe, Grosskopf, Lindgren and Roos (1989), defined their output-oriented Malmquist index as the geometric mean of (3.2.12) and (3.2.14); i.e.,

$$M_o(x^t, y^t, x^{t+1}, y^{t+1}) = \left(\frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^t, y^t)} \right)^{\frac{1}{2}}. \tag{3.2.17}$$

Clearly, this is in the spirit of Fisher (1922) who defined his ideal price index as the geometric mean of the Laspeyres and Paasche indexes.

The productivity index (3.2.17) may be illustrated in a Figure 3.3.

In this figure, two technologies are involved, one for period t and one for period $t+1$. The latter contains the former, i.e., technical progress has occurred. The productivity change for the two input-output vectors (x^t, y^t) and (x^{t+1}, y^{t+1}), based on y-distances in Figure 3.3 is

$$M_o(x^t, y^t, x^{t+1}, y^{t+1}) = \left(\frac{Od}{Oe} \frac{Ob}{Oa} \frac{Oc}{Oa} \frac{Od}{Of} \right)^{\frac{1}{2}} \tag{3.2.18}$$

As we will see in Section (3.2.5), this index can be decomposed into an efficiency and a technical change component, namely

$$M_o(x^t, y^t, x^{t+1}, y^{t+1}) = \left(\frac{Od}{Of} \frac{Ob}{Oa} \right) \left(\frac{Of}{Oe} \frac{Oc}{Ob} \right)^{\frac{1}{2}}, \tag{3.2.19}$$

[7] Caves, Christensen and Diewert (1982) used the geometric mean of (3.2.12) and (3.2.14) to show the relationship between the Törnqvist and the Malmquist productivity indexes.

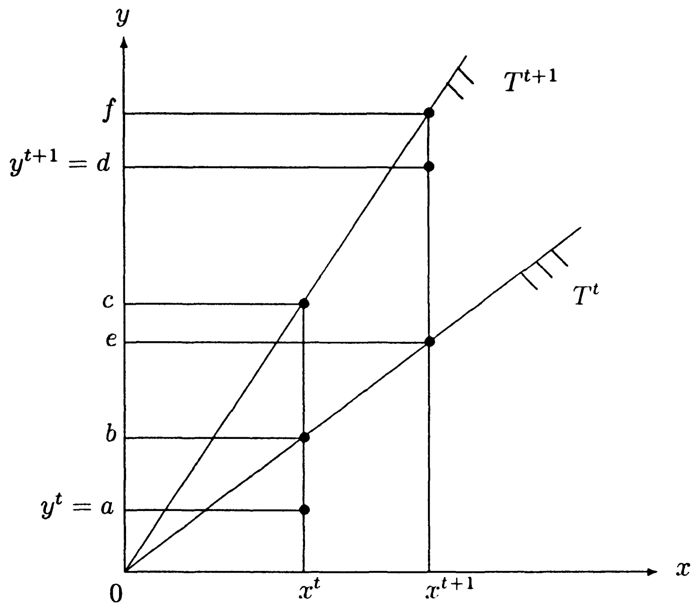

Figure 3.3 The FGLR Output-Oriented Malmquist Productivity Index

where the terms in the first parentheses measure efficiency change and the second term captures technical change.

Diewert (1992) suggested an alternative approach to defining a productivity index using distance functions, namely as ratios of Malmquist output and input quantity indexes. He attributed this idea to Hicks and Moorsteen.[8] The Hicks-Moorsteen approach has been adopted by[9] Bjurek (1994, 1996), Diewert (1993), Grifell-Tatjé and Lovell (1994).

The Malmquist output quantity index at period t is defined as

$$Q_o^t = \frac{D_o^t(x^t, y^{t+1})}{D_o^t(x^t, y^t)} \tag{3.2.20}$$

[8]Diewert (1992, p. 240). See Hicks (1961), Moorsteen (1961) and also Sudit (1995, p. 444).

[9]Bjurek (1994) and Diewert (1993) use geometric means in their adaptations.

and the corresponding input quantity index is

$$Q_i^t = \frac{D_i^t(y^t, x^{t+1})}{D_i^t(y^t, x^t)}. \tag{3.2.21}$$

The Hicks-Moorsteen t-period productivity index is then given by

$$(HM)^t = \frac{D_o^t(x^t, y^{t+1})/D_o^t(x^t, y^t)}{D_i^t(y^t, x^{t+1})/D_i^t(x^t, y^t)} = Q_o^t/Q_i^t. \tag{3.2.22}$$

In order to identify the conditions under which the t-period Hicks-Moorsteen and the t-period Malmquist index coincide, we introduce the notion of inverse homotheticity. The technology is said to exhibit inverse homotheticity if the output distance function takes the form

$$D_o(x, y) = D_o(\overline{x}, y)/F(D_i(\overline{y}, x)), \tag{3.2.23}$$

where F is increasing and $\overline{x}, \overline{y}$ are arbitrarily fixed input and output vectors. Färe and Primont (1995) take these to be the unit vectors.[10] If the technology exhibits constant returns to scale, then F is the identity function.

The following theorem is due to Färe, Grosskopf and Roos (1996).

Theorem: The Malmquist productivity index (3.2.12) equals the Hicks-Moorsteen index (3.2.22) if and only if the technology is inversely homothetic and exhibits constant returns to scale.

Some intuition concerning this result may be given by comparing the (HM)^t index in (3.2.11) and the t-period version of the Malmquist index in (3.2.12). Both use period t as the reference technology. At first blush, it appears that the CCD Malmquist index is equivalent to the numerator of the (HM)^t index, but that is not quite true. The numerator of the (HM)^t index includes $D_o^t(x^t, y^{t+1})$ which evaluates the combination of period t input and period $t+1$ output to the t-period frontier. The CCD index evaluates corresponding data, i.e., input and outputs from the same period. The

[10]Shephard (1970) introduced the idea of inverse homotheticity.

denominator of the $(HM)^t$ index also includes a distance function with 'mixed' data.

3.2.3 Returns to Scale

Berg, Førsund and Jansen (1992)[11] showed that, under constant returns to sale, in the one-input one-output case the Malmquist indexes (2.2.1) and (2.2.3) equal total factor productivity, defined as

$$TFP = \frac{y^{t+1}/x^{t+1}}{y^t/x^t}. \tag{3.2.24}$$

An example by Grifell-Tatjé and Lovell (1995) verifies that the Malmquist indexes (3.2.12) and (3.2.14) are not TFP measures in the sense of (3.2.24) if constant returns to scale is not imposed. This prompted Färe and Grosskopf (1996) to prove

Theorem: Suppose one input is used to produce one output; $TFP = \frac{y^{t+1}/x^{t+1}}{y^t/x^t} = \frac{D_o^\tau(x^{t+1},y^{t+1})}{D_o^\tau(x^t,y^t)}$ $\tau = t$ or $t+1$, if and only if the τ-period technology exhibits constant returns to scale.

This theorem has been used by Färe and Grosskopf (1996) to motivate their use of the constant returns to scale (CRS) technology as the reference technology for the overall Malmquist index and its technical change component. They also show how deviations from CRS may be identified through a scale efficiency term. See also Färe, Grosskopf and Lovell (1994) and Färe, Grosskopf, Norris and Zhang (FGNZ) (1994). This is discussed in more detail in Section 3.2.5.

3.2.4 The Malmquist, Fisher and Törnqvist Productivity Indexes

One of the important achievements of Caves, Christensen and Diewert (1982) was to show that the Törnqvist productivity index could

[11] Berg, Førsund and Jansen (1992, p. S.217) performed their analysis with the input oriented productivity measures. Here we choose to use the output oriented approach.

be derived from the Malmquist index (3.2.17). In the derivation, the authors used the condition that the distance functions are of the translog form. Subsequently, Diewert (1992), also using functional form specifications of the distance functions, proved that both (3.2.12) and (3.2.14) equal the Fisher productivity index. Balk (1993), generalizing a result by Färe and Grosskopf (1992), showed that the Fisher productivity index is approximately equal to the Malmquist index (3.2.17). That proof does not require a specific functional form, but rather relies on a Mahler inequality.[12]

To continue, let us define the Fisher productivity index. This index

$$FP(p^{t+1}, y^{t+1}, p^t, y^t, w^{t+1}, x^{t+1}, w^t, x^t) = \frac{FI_o(p^{t+1}, y^{t+1}, p^t, y^t)}{FI_i(w^{t+1}, x^{t+1}, w^t, x^t)}. \tag{3.2.25}$$

is the ratio of the Fisher Ideal Output Quantity Index,

$$FI_o(p^{t+1}, y^{t+1}, p^t, x^t) = \left(\frac{p^t y^{t+1}}{p^t y^t} \frac{p^{t+1} y^{t+1}}{p^{t+1} y^t}\right)^{\frac{1}{2}} \tag{3.2.26}$$

where $p = (p_1, \ldots, p_M)$ denotes output prices; to the Fisher Input Quantity Index,

$$FI_i(w^{t+1}, x^{t+1}, w^t, x^t) = \left(\frac{w^t x^{t+1}}{w^t x^t} \frac{w^{t+1} x^{t+1}}{w^{t+1} x^t}\right)^{\frac{1}{2}}, \tag{3.2.27}$$

where $w = (w_1, \ldots, w_N)$ denotes input prices. We note that if only one input is used to produce one output, the Fisher productivity index collapses into the total factor productivity measure (3.2.24). If in addition the technology exhibits constant returns to scale, it equals each of the three Malmquist indexes (3.2.12), (3.2.14) and (3.2.17). In general, we can prove the following two theorems.

---

[12] Let $p \in \Re^M_+$ be a vector of output prices, then the revenue function is defined as

$$R(x, p) = \max\{py : D_o(x, y) \leq 1\}.$$

The Mahler (1939) inequality referred to here is

$$R(x, p) D_o(x, y) \geq py, (x, y) \in T.$$

Theorem: The Malmquist output-oriented productivity index (3.2.17) is approximately equal to the Fisher productivity index (3.2.25).

Theorem: If the output distance function for period $\tau = 1,\, t+1$ is of the form $D_o^\tau(x, y) = \sigma^\tau[y\prime Ay(x\prime Cx)^{-1} + \alpha^\tau y\beta^\tau(\frac{1}{x})y\prime B^\tau(\frac{1}{x})]^{\frac{1}{2}}$ where $(1/x) \equiv (1/x_1, \ldots, 1/x_N)$ and "$\prime$" denotes the transpose, and some other conditions given by Diewert (1992, p. 241) hold, then the Fisher productivity index (3.2.25) is equal to each of the Malmquist indexes (3.2.12) and (3.2.14).

The proof of the first theorem is found in Balk (1993) and Färe and Grosskopf (1996), while the proof of the second theorem is found in Diewert (1992). From the comments that (3.2.12) equals (3.2.14) if and only if the output distance function is of the form (3.2.5), it is clear that the functional form of the second theorem has strong implications. In particular it implies that the technology is Hicks output-neutral.

The last two theorems provide strong arguments in favor of using the Fisher productivity index.[13] Actually Diewert (1992, p. 196) argues in its favor: "Our recommended approach to productivity measurement in the general case is the approach outlined in Section 4.4 above, where productivity change was defined as a Fisher quantity index of outputs divided by a Fisher quantity index of inputs;"

The Törnqvist productivity index is defined as the ratio of a Törnqvist output and input quantity index. The output index is

$$TI_o(p^{t+1}, y^{t+1}, p^t, y^t) = \Pi_{m=1}^{M} \left(\frac{y_m^{t+1}}{y_m^t}\right)^{\frac{1}{2}\left(\frac{p_m^{t+1}y_m^{t+1}}{p^{t+1}y^{t+1}} + \frac{p_m^t y_m^t}{p^t y^t}\right)} \qquad (3.2.28)$$

[13] Diewert (1992) also shows from the test results that approach yield arguments in favor of the Fisher productivity index.

and the input index is

$$TI_i(w^{t+1}, x^{t+1}, w^t, x^t) = \Pi_{n=1}^{N} \left(\frac{x_n^{t+1}}{x_n^t} \right)^{\frac{1}{2}\left(\frac{w_n^{t+1} x_n^{t+1}}{w^{t+1} x^{t+1}} + \frac{w_n^t x_n^t}{w^t x^t} \right)}. \tag{3.2.29}$$

Thus the Törnqvist productivity index equals

$$TP(p^{t+1}, y^{t+1}, p^t, y^t, w^{t+1}, x^{t+1}, w^t, x^t) = \frac{TI_o(p^{t+1}, y^{t+1}, p^t, x^t)}{TI_i(w^{t+1}, x^{t+1}, w^t, x^t)}. \tag{3.2.30}$$

Caves, Christensen and Diewert (1992) proved that the geometric mean form of the Malmquist index (3.2.17) equals the Törnqvist index (3.2.30) under certain conditions (see CCD (1982), Theorem 3 for precise details). Some of the conditions worth noting here are that the distance functions are of the translog form[14] (with second order parameters $\alpha_{mm'}$, $\beta_{nm'}$ and $\lambda_{mm'}$ constant across periods and outputs and inputs), optimizing behavior is assumed (revenue maximization and cost minimization), and technology satisfies constant returns to scale (otherwise the Törnqvist index must be adjusted by a scale factor to equal the Malmquist index).

The three theorems of this section show that the Malmquist productivity indexes form the basis for the Fisher and Törnqvist indexes. That is, the Malmquist productivity indexes are more general and include the Fisher and Törnqvist indexes as special cases. The Malmquist indexes also are less 'demanding' in terms of data requirements: no prices or shares are required to compute them, nor is optimizing behavior required. The Fisher and Törnqvist indexes share the advantage of computational simplicity: they can be computed without estimating distance functions.

[14]The distance function $D_o^t(x^t, y^t)$ is said to be of Translog form if $\ln D_o^t(x^t, y^t) = \alpha_0^t + \sum_{n=1}^{N} \beta_n^t \ln x_n^t + \sum_{m=1}^{M} \alpha_m^t \ln y_m^t + \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{n} \beta_{nn'}^t (\ln x_n^t)(\ln x_{n'}^t) + \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \alpha_{mm'}^t (\ln y_n^t)(\ln y_{m'}^t) + \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{nm}^t (\ln x_n^t)(\ln y_m^t)$.

3.2.5 Decompositions of the Malmquist Productivity Index

Nishimizu and Page (1982) decomposed total factor productivity into "technological progress and changes in technical efficiency," (p. 920-921) using a production function and calculus. Subsequently, Färe, Grosskopf, Lindgren and Roos (1994)[15] showed that the Malmquist index (3.2.17) could be decomposed into the same components, namely an efficiency change component[16]

$$\mathrm{EFFCH} = D_o^{t+1}(x^{t+1}, y^{t+1}) / D_o^t(x^t, y^t) \tag{3.2.31}$$

and a technical change component

$$\mathrm{TECH} = \left(\frac{D_o^t(x^{t+1}, y^{t+1})}{D_o^{t+1}(x^{t+1}, y^{t+1})} \frac{D_o^t(x^t, y^t)}{D_o^{t+1}(x^t, y^t)} \right)^{\frac{1}{2}}. \tag{3.2.32}$$

When constant returns to scale is imposed on the reference technologies, the two component measures can be further decomposed. To illustrate these decompositions, some additional notation is needed: $D_o(x, y|C)$ is the output distance function defined on a constant returns to scale technology, and $D_o(x, y|V)$ is the output distance function defined on a variable returns to scale technology.[17] Following Byrnes, Färe and Grosskopf (1984)[18], the output oriented measure of Scale Efficiency is defined by

$$S_o(x, y) = D_o(x, y|V) / D_o(x, y|C). \tag{3.2.33}$$

The constant and variable returns to scale technologies are represented in Figure 3.4. Scale efficiency for (x, y) is measured by the

[15] This paper was first read at the conference on new uses of DEA held at the IC^2 Institute of the University of Texas at Austin on September 27-29, 1989.

[16] Of course, the indexes (3.2.12) and (3.2.14) may also be decomposed into an efficiency change and a technical change component.

[17] In connection with the activity analysis model, "C" means that the intensity variables are nonnegative, while "V" means that in addition they sum to one. See Section 3.2.7.

[18] This measure has it origin in Førsund and Hjalmarsson (1979). For other references, see e.g., Färe, Grosskopf and Lovell (1994), Banker and Thrall (1992), and Førsund and Hjalmarsson (1979, 1987).

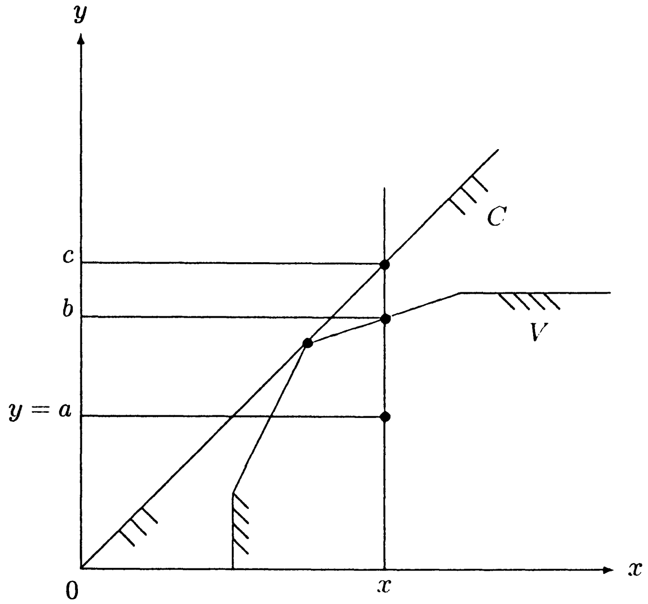

Figure 3.4 Scale Efficiency

distances on the y-axis as

$$S_o(x, y) = (Oa/Ob)/(Oa/Oc) = (Oc/Ob). \tag{3.2.34}$$

Scale efficiency measures deviations from constant returns to scale.[19] We may incorporate the scale measure into the efficiency change component (3.2.31). In this case, we have

$$\text{EFFCH} = \frac{S_o^t(x^t, y^t)}{S_o^{t+1}(x^{t+1}, y^{t+1})} \frac{D_o^{t+1}(x^{t+1}, y^{t+1}|V)}{D_o^t(x^t, y^t|V)}, \tag{3.2.35}$$

i.e., EFFCH consists of scale efficiency change component and an efficiency change component. The latter is now measured relative to the variable returns to scale reference technology. We see scale inefficiency as a short run adjustment problem, thus the inefficiency should be related to the variable returns to scale technology, VRS.

[19]This deviation may be due to increasing returns to scale or (as in the figure) decreasing returns to scale. See Byrnes et al. (1984) or Färe et al. (1994) for how this can be determined.

However, technical change is a long run problem, thus it is measured relative to the constant returns to scale technology, CRS.

The technical change part of the Malmquist index may also be decomposed into component measures. In particular, we may define an input and an output bias term. These together with a magnitude term make up the technical change component.[20]

One way (there are others) to define the Output Biased Technical Change Component, again under constant returns to scale, is

$$\mathrm{OBTECH} = \left(\frac{D_o^t(x^{t+1}, y^{t+1}|C)}{D_o^{t+1}(x^{t+1}, y^{t+1}|C)} \frac{D_o^{t+1}(x^{t+1}, y^t|C)}{D_o^t(x^{t+1}, y^t|C)} \right)^{\frac{1}{2}} \tag{3.2.36}$$

and illustrate it in Figure 3.5.

The two output sets $P^{t+1}(x^{t+1})$ and $P^t(x^{t+1})$ are from different periods but the input vectors coincide. The output vectors y^t and y^{t+1} are both on their respective output isoquants, making $D_o^t(x^{t+1}, y^t|C)$ and $D_o^{t+1}(x^{t+1}, y^{t+1}|C)$ equal one. The bias component then equals,

$$\mathrm{OBTECH} = \left(\frac{Oa}{Ob} \frac{Oc}{Od} \right)^{\frac{1}{2}}. \tag{3.2.37}$$

Färe and Grosskopf (1996) show that OBTECH equals one if and only if the technology is implicit Hicks output neutral,[21] i.e., the output distance function takes the form

$$D_o^t(x^t, y^t) = \acute{D}_o(x^t, y^t)/A(t, x^t) \tag{3.2.38}$$

Input Biased Technical Change can be defined by (again there are alternative ways)

$$\mathrm{IBTECH} = \left(\frac{D_i^t(y^t, x^t|C)}{D_i^{t+1}(y^t, x^t|C)} \frac{D_i^{t+1}(y^t, x^{t+1}|C)}{D_i^t(y^t, x^{t+1}|C)} \right)^{\frac{1}{2}}, \tag{3.2.39}$$

[20] Färe, Grifell-Tatjé, Grosskopf and Lovell (1997) decomposed the t-period Malmquist index (3.2.12). Here we follow Färe and Grosskopf (1996) who decomposed the geometric mean index (3.2.17).

[21] See Chambers and Färe (1994) for details.

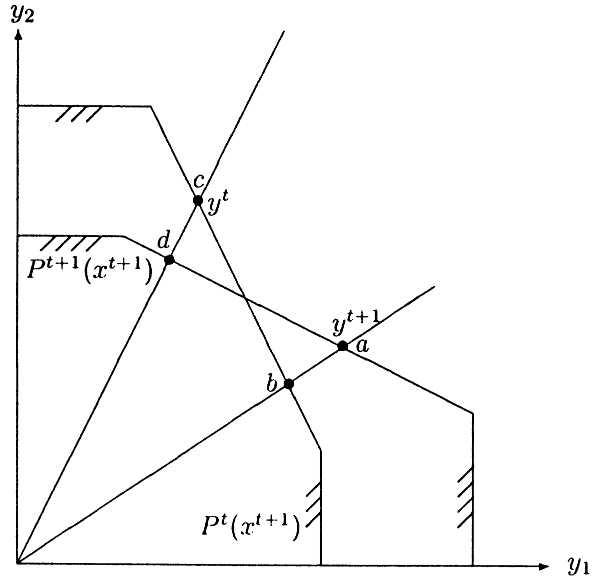

Figure 3.5 The FGLR Output-Oriented Malmquist Productivity Index

and since we have imposed constant returns to scale, (3.2.37) may also be expressed in terms of the output distance functions.[22] We have seen that under constant returns to scale, $D_o(x, y) = 1/D_i(y, x)$.

Finally, the Magnitude Component is defined as

$$\text{MATECH} = D_o^t(x^t, y^t|C)/D_o^{t+1}(x^t, y^t|C). \qquad (3.2.40)$$

It follows that

$$\text{TECH} = \text{OBTECH} \bullet \text{IBTECH} \bullet \text{MATECH}, \qquad (3.2.41)$$

i.e., technical change can be decomposed into three components, accounting for input and output biased technical change and magnitude.

Ray and Desli (1995) suggested an alternative decomposition of the Malmquist index (3.2.17). Their decomposition consists of three

[22] Färe and Grosskopf (1996) showed that IBTECH $= 1$ if and only if the technology is implicit Hicks input neutral, i.e., $D_i^t(y^t, x^t) = B(t, y^t)\hat{D}_i(y^t, x^t)$.

components: technical change, scale efficiency change and efficiency change but with variable rather than constant returns to scale components. In particular of

$$\left(\frac{D_o^t(x^t, y^t|V)}{D_o^{t+1}(x^t, y^t|V)} \frac{D_o^t(x^{t+1}, y^{t+1}|V)}{D_o^{t+1}(x^{t+1}, y^{t+1}|V)}\right)^{\frac{1}{2}} \quad \text{(technical change)} \tag{3.2.42}$$

$$\left(\frac{S_o^t(x^{t+1}, y^{t+1})}{S_o^t(x^t, y^t)} \frac{S_o^{t+1}(x^{t+1}, y^{t+1})}{S_o^{t+1}(x^t, y^t)}\right)^{\frac{1}{2}} \quad \text{(scale efficiency change)} \tag{3.2.43}$$

$$\frac{D_o^{t+1}(x^{t+1}, y^{t+1}|V)}{D_o^t(x^t, y^t|V)} \quad \text{(efficiency change)} \tag{3.2.44}$$

The two components (3.2.42) and (3.2.43) differ from the earlier version of technical change and scale efficiency change. In particular, they contain mixed period distance functions with variable returns to scale, i.e., $D_o^{t+1}(x^t, y^t|V)$ and $D_o^t(x^{t+1}, y^{t+1}|V)$. These distance functions may be difficult to compute, especially if linear programming is used in the calculation.[23] This problem is discussed in Section 3.2.7.

As the following example shows, the scale component (3.2.43) may report values that do not reflect the scale properties of the data. Suppose we have two firms, A and B which use one input to produce a single output.

Firm	(x^t, y^t)	(x^{t+1}, y^{t+1})
A	(6,3)	(5,5)
B	(4,2)	(3,3)

Table 3.1

Based on this sample, both firms are producing at a point of CRS in each period, t and $t+1$. Since both firms were at CRS in both periods, the scale component (3.2.33) registers a value of one, reflecting

[23]This is exemplified in the Ray and Desli (1995) paper by Ireland for which technical change and scale efficiency is not possible to report.

no change in scale efficiency, as it should. The Ray and Desli decomposition gives a scale change value for firm A of $(2/1.66667)^{1/2}$ which is approximately 1.095, suggesting an improvement in scale efficiency, which did not occur.

Moreover, using (3.2.42) to measure technical change will hinder the decomposition into bias and magnitude terms, since the output oriented variable returns to scale distance function need not be reciprocal to the corresponding input oriented distance function. If we multiply the technical change component (3.2.42) with the efficiency change part (3.2.44) then we obtain the Malmquist index (3.2.17) evaluated towards the variable returns to scale technology. Thus the CRS Malmquist index (3.2.17) can be decomposed into a VRS Malmquist index and a scale component (3.2.43).

3.2.6 The Circular Test and Malmquist Productivity Indexes

Recall that if there are three time periods t_1, t_2 and t_3, and if I is an index, then the index satisfies the Circular Test if

$$I(t_1, t_3) = I(t_1, t_2) I(t_2, t_3). \tag{3.2.45}$$

Førsund (1990) and Berg, Førsund and Jansen (1992) pointed out that the Malmquist indexes (3.2.12), (3.2.14) and (3.2.17) do not meet the circular test. This test was embraced by Frisch (1936)[24] and rejected by Fisher (1922). We note that Frisch was only concerned with price indexes, where circularity would be expected. Fisher argued that productivity which has a natural order, namely time, might be expected to be path dependent, i.e., not circular.

The observation by Førsund (1990) and Berg, et al. (1992) led them to introduce a fixed base index, in particular a fixed base[25] t period

[24] See Althin (1995) for a discussion of this and other tests of the Malmquist productivity indexes.

[25] Førsund (1990) and Berg et al. (1992) developed their ideas within the framework of an input oriented index. See also Klein, Schmidt and Yaisawarng (1992).

Malmquist output oriented index defined as

$$M_F = D_o^{t_o}(x^{t+1}, y^{t+1})/D_o^{t_o}(x^t, y^t). \tag{3.2.46}$$

The fixed base is denoted by t_o. Althin (1995) proved that if (3.2.46) is to be independent of the fixed base t_o then (and only then) must the distance function be of the following form:

$$D_o^{t_o}(x^t, y^t) = \hat{A}(t_o)\hat{D}_o(x^t, y^t), \quad \text{for all } t_o. \tag{3.2.47}$$

This is the same form required for the t-period CCD index (3.2.12) to equal, the $t+1$ period CCD index (3.2.14).

Färe and Grosskopf (1996) proved that the Malmquist productivity index (3.2.17) satisfies the circular test if and only if the output distance function is of the form[26]

$$D_o^t(x^t, y^t) = \hat{A}(t)\hat{D}_o(x^t, y^t), \quad \text{for all } t. \tag{3.2.48}$$

Thus again Hicks output neutrality must be imposed.

Recently, Balk and Althin (1996) addressed the circular test issue for the Malmquist productivity index. They created a transitive index that depends on the number of time periods considered, i.e., $t = 1, \ldots, \bar{t}$. Their index is independent of $\bar{t}$ provided the distance functions are of the form (3.2.48).[27]

To conclude, if the circular test is to be satisfied, then the production technology must be neutral in the sense of (3.2.48).

Finally, in a more general framework Aczél (1990) shows that an index $I(t_1, t_2)$ satisfies the circular test if and only if I is multiplicatively separable; i.e.,

$$I(t_1, t_2) = \hat{I}_1(t_1)\hat{I}_2(t_2). \tag{3.2.49}$$

This indicates why the technologies discussed here must satisfy (3.2.48).

[26] Constant returns to scale is used in the Färe and Grosskopf (1996) treatment.

[27] Balk and Althin (1996) also discuss an input oriented measure, i.e., the requirement is that the input distance functions are of the form $D_i^t(y^t, x^t) = \hat{B}(t)\hat{D}_i(y^t, x^t)$.

3.2.7 Computation of the Malmquist Productivity Indexes

There are essentially four different approaches to computing the distance functions that make up the Malmquist productivity indexes. These may be summarized in a table.

	Nonparametric	Parametric
Deterministic	Data Envelopment Analysis	"Aigner-Chu"
Stochastic	Stochastic Activity Analysis	Stochastic Production Frontiers

Table 3.2

The DEA (Data Envelopment Analysis)[28] approach consists of solving a linear programming problem for each firm $k = 1, \ldots, K$ at each period $t = 1, \ldots, \bar{t}$. To fix ideas, suppose the data are given by

$$\{(x^{k,t}, y^{k,t}) : k = 1, \ldots, K, t = 1, \ldots, \bar{t}\}. \tag{3.2.50}$$

A t-period DEA-technology may be formed from the data as

$$\begin{aligned} T^t = \{(x^t, y^t) : \quad & \sum_{k=1}^{K} z_k y_{km}^t \geq y_m^t, m = 1, \ldots, M, && (3.2.51) \\ & \sum_{k=1}^{K} z_k x_{km}^t \leq x_n^t, n = 1, \ldots, N, && (3.2.52) \\ & z_k \geq 0, k = 1, \ldots, K\}. \end{aligned}$$

This is the activity analysis model originated by von Neumann; see Karlin (1959, p. 340). It satisfies constant returns to scale and free disposability of inputs and outputs.[29] Now, distance functions may

[28] The expression Data Envelopment Analysis was coined by Charnes, Cooper and Rhodes (1978).

[29] Inputs are freely disposable if $(x, y) \in T$ and $x\prime \geq x$ imply $(x\prime, y) \in T$. Outputs are freely disposable if $(x, y) \in T$ and $y\prime \leq y$ imply $(x, y\prime) \in T$.

be computed relative to the reference technology T^t. For example,

$$\begin{aligned}(D_o(x^{k',t}, y^{k',t}))^{-1} &= \max \theta \\ &\text{s.t. } \sum_{k=1}^{K} z_k y_{km}^t \geq \theta y_{k'm}^t, m = 1, \ldots, M, \\ &\sum_{k=1}^{K} z_k x_{kn}^t \leq x_{k'n}^t, n = 1, \ldots, N, \\ &z_k \geq 0, k = 1, \ldots, K.\end{aligned} \tag{3.2.53}$$

Here the technology and the observation k' are from the same period. In the case that they are not, e.g., in

$$\begin{aligned}(D_o^t(x^{k',t+1}, y^{k',t+1}))^{-1} &= \max \theta \\ &\text{s.t. } \sum_{k=1}^{K} z_k y_{km}^t \geq \theta y_{k'm}^{t+1}, m = 1, \ldots, M, \\ &\sum_{k=1}^{K} z_k x_{kn}^t \leq x_{k'n}^{t+1}, n = 1, \ldots, N, \\ &z_k \geq 0, k = 1, \ldots, K,\end{aligned} \tag{3.2.54}$$

a maximum may not exist. (See Färe, Grosskopf and Roos (1995) and references therein.) However if all inputs and outputs are positive, this problem does not occur under CRS.[30]

The second deterministic method, which we refer to as the Aigner-Chu (1968) method, works by parameterizing the distance function and then computing its parameters using linear programming. For example, if the output distance function is assumed to be of translog form, then the following problem must be solved.

$$\begin{aligned}&\text{minimize } \sum_{k=1}^{K} [\ln D_o^t(x^{k,t}, y^{k,t}) - \ln 1] \\ &\text{subject to} \\ &\quad \text{i)} \quad \ln D_o^t(x^{k,t}, y^{k,t}) \leq 0, k = 1, \ldots, K, t = 1, \ldots, \bar{t}\end{aligned} \tag{3.2.55}$$

[30] If variable returns to scale is imposed, i.e., the sum of the intensity variables is one ($\sum_{k=1}^{K} z_k = 1$), the maximum may not exist even if all inputs and outputs are positive, see Färe, Grosskopf and Roos (1995).

$$\text{ii)} \quad \sum_{m=1}^{M} \alpha_m^t = 1, t = 1, \ldots, T$$

$$\text{iii)} \quad \sum_{m'=1}^{M} \alpha_{mm'}^t = \sum_{m=1}^{M} \gamma_{nm}^t = 0, m = 1, \ldots, M,$$
$$n = 1, \ldots, N, t = 1, \ldots, \bar{t}$$

$$\text{iv)} \quad \alpha_{mm'}^t = \alpha_{m'm}^t, m = 1, \ldots, M, m' = 1, \ldots, M,$$
$$t = 1, \ldots, T$$
$$\beta_{nn'}^t = \beta_{n'n}^t, n = 1, \ldots, N, n' = 1, \ldots, N,$$
$$t = 1, \ldots, \bar{t}$$

where

$$\begin{aligned}
\ln D_o^t(x^{k,t}, y^{k,t}) &= \alpha_o^t + \sum_{n=1}^{N} \beta_n^t \ln x_n^{k,t} + \sum_{m=1}^{M} \alpha_m^t \ln y_m^{k,t} \\
&+ \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \beta_{nn'}^t \ln x_n^{k,t} \ln x_{n'}^{k,t} \\
&+ \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \alpha_{mm'}^t \ln y_{m'}^{k,t} \ln y_{m'}^{k,t} \ln y_m^{k,t} \\
&+ \sum_{n} \sum_{M} \gamma_{nm}^t \ln x_n^{k,t} \ln y_m^{k,t}.
\end{aligned}$$

In (3.2.55), one set of parameters α^t, β^t, γ^t is calculated for each period $t = 1, \ldots, \bar{t}$. The corresponding output distance function is assumed to hold for each period $t = 1, \ldots, \bar{t}$. The corresponding output distance function is assumed to hold for each observation $k = 1, \ldots, K$, and specific values are obtained by introducing the data say $(x^{k,t}, y^{k,t})$ into the function with the computed parameters. Alternatively, one could also minimize over t, and include trend terms, as in Nishimizu and Page (1982).

With the exception of using chance constraints, the stochastic activity analysis model has not played a large role as of yet in the computation of Malmquist productivity indexes. One may also compute distance functions as stochastic frontier functions.

It is quite interesting to note that the four approaches that are used to compute distance functions can all be derived from the stochastic production correspondence. Let Ω denote the states of the world; then, following Krug (1976), a stochastic output correspondence may be defined as

$$\mathcal{P}(x,w) = \{y : x \text{ can produce } y, \text{ given the state } w\}, w \in \Omega. \tag{3.2.56}$$

The stochastic output distance function is then derived from $\mathcal{P}(x,w)$ as

$$D_o(w,x,y) = \sup\{\theta : (y/\theta) \in \mathcal{P}(x,w)\}. \tag{3.2.57}$$

Now if $D_o(w,x,y)$ is constant with respect to $w \in \Omega$, the usual distance function is obtained.

3.3 EMPIRICAL APPLICATIONS

The Malmquist productivity indexes have been used in a variety of studies. These studies include agriculture, airlines, banking, electric utilities, insurance companies, and public sectors, as well as country comparisons of productivity.

We start our empirical part of the survey with the public sector studies, since it was there that DEA was first used in computing Malmquist productivity changes.

3.3.1 Public Sector Studies

In the first paper using the nonparametric deterministic approach, or DEA, to compute the Malmquist productivity index, Färe, Grosskopf, Lindgren and Roos (1994) analyzed the Swedish hospital sector. In that industry there are no prices that reflect scarcity and hence the Törnqvist and Fisher productivity indexes cannot be computed and are inappropriate. The authors used the output oriented model (3.2.17) with nonincreasing returns to scale imposed i.e., the intensity variables z_k sum to less than or equal to one. For the time period 1970 to 1985, they compute the Malmquist index and its two component measures, technical change and efficiency change. Their results indicate considerable variation in efficiency change among the 17 hospitals in their sample, and the technical change component showed both progress and regress. This study was extended by Färe, Grosskopf and Roos (1994) to include the years 1970-1992 and to employ constant returns to scale for the reference technology (3.2.17). Again the results are mixed but the overall conclusion is that there has been a decline in productivity over that time period.

Burgess and Wilson (1995) undertook a study of types of U.S. hospital ownership and productivity growth. They identify four ownership types: 1) VA hospitals (132 observations), 2) state and local government enterprises, non-VA hospitals (224), 3) nonprofit organizations (1021) and 4) for-profit firms (168) and used data for the time period 1985-1988.[31] Burgess and Wilson use both an input and an output oriented nonparametric formulation of (3.2.17), and they impose variable returns to scale, i.e., the intensity variables z_k sum to one. They found that changes in efficiency are small both over time and groups, so that productivity change is mainly explained by technical change. On average, over all groups there has been technical regress, in particular during 1987/1988.[32]

Magnussen (1994) studies 46 Norwegian hospitals during the time period 1988-1991, with four outputs and three inputs. Magnussen uses the models (3.2.12), (3.2.14), (3.2.18) and two fixed-base indexes (3.2.17).[33] Not surprisingly, he finds large differences among the productivity measures. "When asked whether productivity has improved or worsened over the period 1988 to 1991, the answer improved by 10% or declined by 10% depending on how you see it is hardly convincing." (p. 145). The average over the years and

[31] A special feature of their paper is that they, following Wilson (1995), check the data for outliers.

[32] Burgess and Wilson report 95 percent confidence intervals obtained by bootstrap the geometric means along the lines of Atkinson and Wilson (1995).

[33] Productivity is also measured using sequential frontiers, see Tulkens and Vanden Eeckaut (1995). This model sequentially adds data to the frontier, e.g., the 1989 technology consists of the data from 1988 and 1989.

hospitals for the geometric mean formulation (2.2.6) show 0.998; i.e., there has been no progress. This outcome is consistent with the Swedish studies.

Färe, Grosskopf, Lindgren and Poullier (1997) made an international comparison of productivity growth in health care delivery. They used the Malmquist model (3.2.17) with constant returns to scale to calculate productivity changes for 19 OECD countries over the period 1974-1989. They use two types of output measures. In their first model, outputs are numbers of hospital beddays and inpatient admissions (discharges). This model focuses on the production of intermediate hospital outputs (the intermediate model). In their second model, they focus on the improvements in the health care status of the population. In this outcome model, outputs include the reciprocal of the infant mortality rate and women's life expectancy at age 40.

Bjurek and Hjalmarsson (1995) and Bjurek, Kjulin and Gustafsson (1992) discuss productivity growth in two Swedish public sectors. Bjurek and Hjalmarsson study the public social insurance service and Bjurek et.al., study public day care centers.

Bjurek and Hjalmarsson specify a parametric factor requirement function and use the Aigner-Chu approach to compute the parameters in their function. They use the results to compute a fixed based Malmquist productivity index and its component measures of efficiency change (or catching-up) and technical change. Over the period 1974 to 1987 (with base 1974), there has been a decline in productivity. The decline is attributed to negative technical change. To quote: "The main conclusion of this analysis is that productivity in the sector has decreased considerably over a fairly long period." (p. 457).

Their analysis include 194 day care centers and they use a nonparametric DEA approach to compute productivity index, which over the two years 1988 and 1989 show a 3% decrease.

> "The Malmquist index can be calculated on the basis of one frontier for one particular year or as a geometric mean based on two or more frontiers, one for each year. The choice is not obvious, and the different alternatives all have advantages and disadvantages. We calculated Malmquist indexes based on the specification with one common frontier for both years. Creating both years as one cross section and calculating the indexes based on one frontier, i.e., assuming that the "production technology" is unchanged, are well motivated in this case. Here the focus is not primarly on shifts in the frontiers, the "technical-change" effect, and the Malmquist index is, in this case, reduced to the "catching-up" effect."

Bjurek, Kjulin and Gustafsson (1992, p. S181).

In Sweden, the retail trade of pharmaceutical productions has been the responsibility of the public monopoly Apoteksbolaget since 1971. Färe, Grosskopf, Lindgren and Roos (1992) used the input oriented version of the Malmquist Index (3.2.17) to compute productivity and its components for 42 Swedish group pharmacies during 1980-1989. In Färe, Grosskopf and Roos (1995:a), the earlier productivity index was extended by incorporating attributes into the reference technology. The attributes were used to measure the service quality of each pharmacy, and in particular a quality change component was extracted from the productivity index. The (1995:a) paper was extended by Färe, Grosskopf and Roos (1996) by explicitly introducing a utility function into the Malmquist productivity index. In particular, outputs were evaluated through a utility function and a utility change component could be extracted from the productivity index. Thus consumer satisfaction was explicitly recognized in the productivity index. See also the paper by Norlander and Roos in this volume.

The fixed based Malmquist index (3.2.46) has been used in two Norwegian studies of ferries for the years 1984-1988. Kittelsen and

Førsund (1992) use the same index in their study of Norwegian district courts, during 1983-1988.

Førsund (1993) uses a nonparametric DEA model with constant returns to scale and decomposes productity change into efficiency change (catching-up) and technical change. "The main impression is one of productivity decline due to the average unit gradually falling behind the frontier." (p. 366)

Kittelsen and Førsund (1992) also using a nonparametric constant returns to scale model find: "Total productivity is falling for the first and last period, and in between increases. The improvement for the whole period has been rather weak at about 6% , with 4% due to catching-up and 2% due to technology shift." (p. 295)

The Norwegian road sector is studied by J. Odeck (1993). Chapter three of the dissertation analyses the 67 stations during 1989-1991 that make up the Motor Vehicle Inspectorate. Odeck adopted the same model as in the other two Norwegian studies, i.e., a fixed based nonparametric Malmquist index, and he finds some productivity progress using 1989 as the base year.

Taskin and Zaim (1995) use the nonparametric Malmquist model (3.2.17) to study the public enterprise sector in Turkey during 1974-1991. The authors decompose the model into technical and efficiency changes and decompose the latter into scale and pure technical efficiency changes. Their results show that the growth in the public sector was 14% on average over the period and 37% for the private sector. The major reason for the growth is technical change while there has been a decline in the efficiency component. In both sectors, the decline in the efficiency component is due to the scale component. Firms have not adjusted to the optimal scale fast enough.

3.3.2 The Banking Industry

This section summarizes some of the work which uses Malmquist productivity indexes to analyze the performance of the banking industry. Most of the studies analyze the performance of banks within one country, although a few have undertaken to make international comparisons. The studies use several forms of the Malmquist index including the fixed base model similar to (3.2.46), Malmquist indexes based on free disposal hull technology with variations on how to incorporate time (intertemporal, sequential, or contemporaneous), the original CCD index as in (3.2.12) and (3.2.14) under various returns to scale, to mention a few. Decompositions include the simple two components: efficiency change and technical change. Alternatives also include a scale component or decompose technical change in magnitude and bias terms. Although most of these papers eschew statistical tests, there are some notable exceptions which compute confidence intervals by asset size class. Most of the papers were motivated by recent deregulation in this industry which has occurred in many countries.

Tulkens and Malnero (1996) use Malmquist productivity indexes of the form in (3.2.17) but based on a free disposal hull definition of the underlying distance functions to analyze the productivity (including efficiency change and technical change components) of 663 branches of one bank in Belgium over an 11 month period in 1987.[34] One of the contributions of this paper was to compare three different ways of defining the reference technology when computing the indexes. These three variations include a contemporaneous frontier, a sequential frontier and an intertemporal frontier.[35] The contemporaneous frontier is formed by computing efficiency separately in each period. The sequential frontier is formed by adding data according to the march of time; unlike the windows technique, the reference set becomes larger over time. In the intertemporal case,

[34] This adaptation of the FGLR decomposition was developed first in Tulkens and Vanden Eeckaut (1995).

[35] These three approaches to analyzing performance over time were first proposed in Tulkens (1986).

all the data from all time periods is pooled. The authors provide results for the three approaches, and show how one may compare results across regions and size classes.

Berg, Førsund, and Jansen (1992) compute input-based productivity change for Norwegian banks during the 1980's when the banking industry was deregulated. The form they choose is the original CCD single ratio form; however, they did employ the efficiency change and technical change decomposition following Färe, Grosskopf, Lindgren and Roos. One of the issues raised in this paper is the choice of a reference period. The authors argue for using a fixed base period, with the rationale being that the indexes will then satisfy the circular test. The compute an index of the form in (3.2.46). Their data set is a balanced panel of the 152 banks which survived over the 1980-89 period. Their model includes three outputs (value of short-term loans, long-term loans, and deposits) and two inputs (hours and a residual 'materials' input). Their results from this model suggest regress in the earlier years, and progress in the later years of the sample on average. They also experiment with including loan losses in the output vector.

In a contribution to the special issue on banking in the *Journal of Banking and Finance*, Berg, Førsund, Hjalmarsson, and Suominen (1993) use the Malmquist index to compute multilateral productivity indexes for Nordic banks in the year 1990.[36] They use the single ratio form from CCD, i.e., (3.2.12). Their specification includes total value of loans, total value of deposits and number of branches as outputs. Inputs include labor (hours) and capital (book value). They compute their productivity indexes relative to both CRS and VRS technologies and decompose productivity into a relative efficiency and relative technology component. Their findings suggest that large Swedish banks are the most likely to expand into the Nordic banking market.

Another study which uses the same general technique, i.e., a multi-

[36] Their data include 503 Finnish banks, 150 Norwegian banks and 126 Swedish banks.

lateral comparison based on the single ratio CCD Malmquist index, is by Pastor, Pérez and Quesada (1994). They also undertake an international comparison, but a broader one. They have data on commercial banks operating in 1992 from Austria, Belgium, France, Germany, Italy, Spain, the United Kingdom and the United States. They specify several models: the first includes two inputs (personnel expenditures and non-interest expenses) and three outputs (loans, other assets and deposits). They compute two other models: one in which employment is substituted for personnel expenditures and number of branches is included as an environmental characteristic, and the second which also includes a capital ratio constraint to capture exposure to risk. They impose CRS for computation of their multilateral indexes. An interesting feature of their study is the presentation of results; they include a 'medium bank' (presumably a hypothetical bank constructed at the mean of the data for each country), an unweighted average and a weighted average. They find considerable evidence of inefficiency and variation in productivity.

In a series of four papers, Grifell-Tatjé and Lovell analyze the productive performance of Spanish banks. Three of these focus solely on savings banks and employ the same panel of data from 1986-91. In these papers, the authors use the number of loans, the number of checking accounts and the number of savings accounts as their output measures.[37] Inputs included employment, expenditure on materials, and expenditure on capital. These three papers differ in the way in which productivity is computed. In the 'Scale Economics ... ' paper, the authors compute what they refer to as the 'generalized Malmquist index', which is a version of the Hicks-Moorsteen index (3.2.22).[38] In 'A DEA-Based Analysis of Productivity Change ... ', they use what appears to be the same model, but without the scale term.[39] The third paper using the

[37] In 'Deregulation and Productivity Decline: The Case of Spanish Savings Banks,' they also consider an augmented model which includes the number of bank branches as an output.

[38] Their scale efficiency change is not defined as in Färe, Grosskopf and Lovell (1994), nor as in (2.5.3). It includes a distance function term with inputs and outputs from different periods.

[39] Although the models appear to be the same, the results for the productivity, efficiency

savings bank data set, 'Deregulation and Productivity Decline ...', uses the geometric mean formulation of the Malmquist index, with CRS technology, and decomposition into technical change and efficiency change, i.e., the model from (3.2.17) with the decomposition in (3.2.31) and (3.2.32). In all three papers, they find pervasive declines in productivity.

The fourth paper, 'Sources of Productivity Growth ...', includes commercial banks as well as savings banks and extends the panel to 1993. The authors also change the they measure inputs and outputs: instead of numbers of accounts, they use values for outputs. Personnel expenditures replace labor, and material and capital expenditures are combined. The model reverts to the CCD type in (3.2.12) computed relative to VRS (with decomposition into efficiency change, technical change, and bias) which they augment with their version of change in scale efficiency. They begin by computing productivity for the commercial and savings banks separately, then, following an idea used in Charnes, Cooper and Rhodes (1978), they correct the data for inefficiency and then pool the savings and commercial banks to identify gains from institutional structure. Interestingly, all of their results suggest that there has been positive productivity growth, which is in marked contrast to their earlier results.

Fukuyama (1995:a) computes an input-based Malmquist productivity index using the FGLR model in (3.2.17) for Japanese banks over the 1989-1991 period, which covers the time period of the 'collapse of the speculative bubble' in Japan: the Nikkei index fell from 39 thousand yen in 1989 to 14 thousand by 1992. His sample includes a variety of types and sizes of banks. His model includes three inputs and revenues from loans and revenue from investment activities as the two outputs.[40] He finds a uniform decline in productivity between 1989 and 1991, with city banks experiencing the largest declines.

change and technical change terms were slightly different in these two papers.

[40]The author notes that interest rates are fairly uniform in Japan.

Bauer, Berger, Ferrier and Humphrey (1995) compute fixed period output-based Malmquist productivity indexes (3.2.46) for a panel of data covering 408 large U.S. banks over the 1977-1988 period. The authors decompose their productivity measure into technical change, pure efficiency change, and scale change. Their specification includes four outputs (value of demand deposits, real estate loans, commercial and industrial loans, and installment loans) and four inputs (labor, physical capital, small time and savings deposits, and purchased funds). The motivation for their study was to analyze the effect of deposit interest rate deregulation and the introduction of new types of consumer accounts in the 1970's. They find a decline in performance in the beginning of their time period, but improvements in the later years. These appear to be due mainly to technological change.

In a pair of studies, Wheelock and Wilson (1994, 1995) compute Malmquist output productivity indexes for all US commercial banks (with complete data) for the 1984-1993 period. They use the FGLR specification from (3.2.17), but impose variable returns to scale[41]. Their model includes three inputs (physical capital, labor and purchased funds), and five outputs (real estate loans, commercial and industrial loans, consumer loans, all other loans, and total demand deposits). Data are from the FEIC Reports of Condition and Income, and yield 11,387 observations in 1993 to a maximum of 14,108 observations in 1985. One of the innovations in their work is to provide tests of statistical significance of changes in average productivity and its components by constructing confidence intervals by asset size class. They find that banks of all size classes experience declines in technical efficiency between 1984 and 1993. On the other hand, banks with assets greater than 300 million improve in terms of productivity; those with less than or equal to 300 million decline in productivity. Their results suggest that a few banks are shifting the frontier, implying that there are large productivity gains to be

[41]Note that since they used variable returns to scale, their productivity results will not be consistent with the notion of technical change as change in highest average product, and productivity change as change in observed average product.

exploited in the future by those that have lagged behind.

Perhaps the most ambitious of the banking studies is that by Weber and Devaney (1996). They compute the FGLR type Malmquist index as in (3.2.17) along with the full decomposition of technical change in (3.2.41). They go on to enhance this decomposition to include a regulatory component. Their goal was to determine whether risk based capital standards introduced in 1991 imposed a binding constraint on productivity growth and whether that regulation induced chanes in the mix of outputs. In particular, they wanted to know whether regulation induced the decline in real estate lending between 1990 and 1993. In order to isolate the regulatory effect, they include models with and without explicit risk based capital constraints.[42] In order to assess changes in mix, they also computed input and outputs biased technical change based on earlier work by Färe and Grosskopf (1996).[43] Weber and Devaney augment the analysis of bias by explicitly identifying 'using' and 'saving' factors. Their results suggest that risk based capital constraints impose a loss, and that biases in technical change occurring over this period favor the riskiest assets category.

Devaney and Weber (1995) also compute Malmquist output based productivity a la (3.2.17) for all rural US banks in 1990, 1992 and 1993. Their specification includes four outputs (securities, real estate loans, personal loans, and commercial loans) and three inputs (labor, capital and deposits). The goal of their study is to test the efficiency market hypothesis for rural banks. Over the 1988-1993 period there was an increase in concentration in the rural banking sector. They wished to determine whether that increase in concentration was accompanied by an improvement in performance as measured by productivity change.[44] They find evidence that pro-

[42]In order to include risk based measures of capital, the authors could only include banks with over 1 billion dollars in assets in their sample.

[43]Later, Färe, Grifell-Tatjé, Grosskopf and Lovell (1997) published a modified version of the earlier work.

[44]They also investigated whether 'adverse selection' contributed to rural credit market failure in the Lower Mississippi Delta Region. They use allocative efficiency as their measure of market failure and use a Tobit analysis to identify the sources of that inefficiency.

ductivity growth was higher in areas in which concentration was falling, i.e., their results are consistent with a market power story rather than with an efficient market story.

3.3.3 Productivity in the Agricultural Sector

Thirtle, Hadley and Townsend (1994) compute input-based Malmquist productivity indexes at the country level for agriculture in sub-Saharan countries (Burking Faso, Burundi, Cameroon, Congo, Ivory Coast, Central African Republic, Ethiopia, Ghana, Kenya, Mali, Malawi, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, Sudan, Tanzania, Togo, Zaire, Zambia and Zimbabwe) over the period 1971-1986. One of their goals is to explain differences in productivity in a second stage regression analysis. They assume that traditional inputs, i.e., land, labor and livestock, are used to produce aggregate agricultural output. In the second stage regression analysis, they include variables to reflect modern inputs and infrastructure investments. To capture policies that lead to adoption of new technology they use R& D expenditures, extension and education. They use a real protection coefficient to capture agricultural price policies, and a weather index. They find that the policy variables explain about 11% of the variation in productivity, but even more important seems to be population pressure. They find that productivity growth is small, but generally positive.

In a series of three papers, Loren Tauer (1994, 1995, 1996) assesses the productivity of New York dairy farms. In the first paper, the author computes output-based Malmquist productivity and decomposes it into technical change and efficiency change as in FGLR. The data consists of a balanced panel of 49 dairy farms over the 1977-1987 period. An interesting feature of his work is that he integrates a chance constraint of his output variable in the (non-linear) programming problem he solves for the distance function. In comparing these results to the more traditional model, he finds that the productivity change measure was highly correlated, but the technical change and efficiency change components had much lower correlations across the two models.

In Tauer (1995) the author uses Malmquist technical change estimates to correct for technical change in order to apply the weak axiom of profit maximization and cost minimization to the same data set as above. The third paper also analyzes New York dairy farms, but a 'wider' panel (70 farms) and more recent period 1985-1993. The author computes a Malmquist output-based productivity index and uses the enhanced decomposition of technical change following Färe, Grifell-Tatjé, Grosskopf and Lovell (1997). He finds that productivity increased with an average of over 2% per year over this period. However, as he points out, for 25% of the sample, increases in productivity were insufficient to offset the unfavorable changes in prices over this period.

In another application to dairy farming, Turk, Piesse and Thirtle (1995) assess the performance of four co-operative and twelve private dairy farms in the Yugoslav republic of Slovenia from 1974-1990. They compute Malmquist input-based productivity indexes based on (3.2.17) and decompose productivity into technical change, efficiency change and change in scale efficiency. The last component ends up being critical in explaining differences in productivity across the two ownership forms: the cooperatives are much larger (and scale efficient) than the private farms. Based on the other components, however, the private farms show more impressive improvement over this time period.

Ferrantino and Ferrier (1996) use Malmquist output-based productivity measures of the fixed base, simple ratio model used by Førsund to assess the performance of the Indian sugar industry. They employ the decomposition into technical change, efficiency change and scale efficiency change. Their data set includes 122 Indian sugar factories operating over five growing seasons between 1981 and 1986. These factories produce two outputs – sugar and molasses – using eight inputs (permanent workers, temporary workers, agricultural sucros, chürshers, boiling and generating capacities and the number of days in the crushing season). During 1983-1985, there was a large drop in sugar cane production which ultimately resulted in measured technical regress. The authors also identify the government's price controls over refined sugar and not over 'gur' sugar as a contributor to this decline in performance. Another interesting feature of this paper is that the authors compute confidence intervals for the mean values of their indexes following Wheelock and Wilson (1994).

Another paper which directly addresses the issue of productivity decline is Piesse, Thirtle and van Zyl (1996), who compute Malmquist productivity for maize production on farms in the South African homelands in 1991 and 1992-1993 which was a drought year. Their goal is to assess the effect of the drought on these farms. They compute input-based Malmquist productivity and decompose it into efficiency change, technical change and scale efficiency change for 174 farms. Their variables include one output (maize production) and four inputs (labor, land, seed and fertilizer). Interestingly, they find that the region that was most 'advanced' (in terms of use of modern inputs and techniques) had the biggest decline in productivity during the drought. They attribute this to the greater risk associated with modern input use. In particular, the use of hybrid seed is highly correlated with productivity decline.

3.3.4 Country Studies

Domazlicky and Weber (1997) compute Malmquist output-based productivity for the US over the 1997-86 period, where the unit of observation is the state. They employ the enhanced decomposition of (VRS) efficiency change, technical change and scale efficiency change. Instead of focusing on manufacturing, they instead employ two outputs: private and public gross state product, and four inputs: private and public labor and private and public capital. The authors find very little productivity growth overall (less than one-half of one percent per year based on a weighted average), but considerable regional variation. Almost all productivity growth is due to technical change, which is also positively and significantly related to output growth as expected. The authors also find that initial levels of efficiency are negatively correlated with productivity growth, suggesting a catching-up form of convergence in productivity.

Using similar data, but a longer time period, Boisso, Grosskopf and Hayes (1996) also compute output-based Malmquist productivity decomposed into technical change and efficiency change. Again, the unit of observaiton is the state; the model includes gross state product as output, and labor and public and private capital as inputs. Although the authors are also interested in regional variations in productivity, one of the issues they address is whether government activity hinders or enhances productivity. Earlier (controversial) work found that public infrastructure capital had a strong positive impact on output and productivity growth. It has been argued that this effect may be amplified by possible spillover effects of that capital on productivity across state boundaries. On the other hand, a large public sector is generally thought to be inefficient and therefore detrimental to productivity growth. The authors use a random effects model to test these alternative hypotheses. They use their productivity and component measures as dependent variables, and include various control variables, including business cycle variables, regional variables, controls for service/manufacturing mix, as well as variables to capture infrastructure, spillover and size of public sector effects. They find a minor positive impact of public capital on productivity and a negative effect of public sector share of output on productivity, but a positive effect on technical change.

In another country study, Ochoro and Hjalmarsson (1995) compute fixed base single ratio Malmquist productivity indexes using firm level data in four manufacturing groups in Kenya between 1993 and 1994. One of the interesting features of their data is the wide variation in size and performance of the firms in their sample, which yield wide variation in performance as expected. They conclude that there are very large potentials for productivity improvements;

but, due to the skewed size and performance distribution, lack of competitive pressure may prevent these from being realized.

Ferrier, Klinedinst, and Linvill (1996) undertake an analysis of productivity for a sample of Yugoslav enterprises for the 1975-79 period. They compute Malmquist output-oriented productivity and decompose it into efficiency change, technical change and scale effects. The unique organizational nature of these enterprises, especially after the 1974 reforms which decentralized control and subdivided into more logical units, motivates their study. The authors compute a fixed base productivity index and find evidence of falling productivity in their sample over this period. They find that regression analysis of various policy/structural variables (market share, capital intensity, joint ventures with foreign firms, share of output exported) was generally significant based on single year performance, but generally not significant for productivity change or its components.

In a pair of papers, Färe, Grosskopf and Lee (1995, 1996) analyze productivity in Taiwan manufacturing. In the first of these papers, the authors compute Malmquist output-based productivity and decompose into technical change and efficiency change for four manufacturing industry groupings during the period 1978-89. The four industries are pooled in the computation. The authors find that productivity improvements are almost solely due to technical progress, and that progress was greater during the liberalization period starting in 1984. Their second paper uses much more disaggregated data for a longer time period: there are 16 manufacturing industries covered from 1978-1992. The authors again compute Malmquist output productivity, but this time decompose their technical change component into output bias, input bias and magnitude of technical change. They find that since the liberalization period there has been a tendency toward capital-using input bias.

Färe, Grosskopf, Norris and Zhang (1994) demonstrated how Malmquist output-based productivity measures could be decomposed into technical change, efficiency change and scale effects, and applied this technique to the analysis of productivity growth and convergence in a sample of OECD countries over the 1979-1988 period. They find that almost half of the relatively high productivity growth observed in Japan was due to catching up rather than innovation, whereas the U.S., which had lower productivity growth overall, was shifting the frontier over this time period for this sample. Perhaps the main contribution of this paper was the demonstration of the technique and connection to the convergence literature.

Perelman (1995) provides an international productivity comparison for a sample of OECD countries over the 1970-1987 time period. The author computes output-based Malmquist productivity (and technical change and efficiency change components) as well as a parametric frontier analog. He also computes productivity separately for eight manufacturing sectors. In order to address the issue of the effect of R&D on productivity growth, the author uses weighted OLS to regress various explanatory variables on productivity growth and its components. These include a ratio of lagged R&D expenditures to GDP, lagged technical efficiency (to capture the catching up process), the ratio of imports and exports to total value of production (to capture international competitiveness), the ratio of new investments to capital, and lagged output growth (to capture exogenous growth of demand). He finds that R&D is positively associated with innovation, and that capital formation and output growth positively affect efficiency change and productivity growth.

In another international comparison, Taskin and Zaim (1995) compute Malmquist output-oriented productivity for a sample of both developed and less developed countries over the 1975-1990 period. They use the decomposition into technical change, efficiency change and scale change to address several issues in the convergence literature: whether efficiency change is faster in 'follower' countries, whether developed countries are producing at increasing returns to scale (as predicted by some variations of the endogenous growth models), and whether increased efficiency can be linked to specific policy environments, such as openness to trade. They do find that efficiency change is much higher for the LDCs, but not enough to compensate for the higher rates of technical change in the developed countries in their sample. To address the issue of openness to trade, the authors use the World Bank classification of degree of openness to trade (available for LDCs) and group their LDCs by these classifications. They find that the more open groups have achieved greater productivity gains.

Gouyette and Perelman (1995) use Malmquist productivity indexes to test for convergence among a sample of 13 OECD countries over the 1970 to 1988 period. Using the geometric form of the output based index decomposed into technical change and efficiency change, they find that although productivity growth rates are low in the service sector, they converge, in contrast to the manufacturing sectors. In a second stage analysis, they regress productivity on initial efficiency levels and growth in the capital labor ratio. The coefficient on the efficiency level is a test for 'catching up'. Interestingly, they find that capital intensity is positively related to productivity in manufacturing, but negatively related in the service industry.

Førsund (1996) uses a fixed based Malmquist index as in Berg, Førsund and Jansen (1992) to study a set of twelve Norwegian establishments during the time period 1976-1988. The overall impresssion is that productivity has declined, using 1976 as the base year. When 1988 is chosen as the base, productivity growth declines for the early periods. Førsund finds that positive productivity growth corresponds with positive output growth, i.e., Verdoorn's Law applies.

Vega (1994) includes a study of productivity for a sample of cement plants in Peru in his dissertation. He computes the single ratio version of input based Malmquist productivity decomposed into technical change and efficiency change. Using a four input, one output specification, he finds positive or constant productivity growth for most of the 6 plants over the 1970-1986 period.

The Peruvian cement industry has been the subject of a number of studies. Cabezas Vega and Veiderpass (1994) compute the single ratio version, like (3.2.12), of the input based Malmquist productivity index, decomposed into technical change and efficiency change. Using a four input one output specification, they find positive or constant productivity growth for most of the plans over the 1970-1986 period. Since the Peruvian (cement) industry has a history of successive, politically motivated, changes betwen private and public ownership, Cabezas Vega and Veiderpass (1992) calculate the input based Malmquist index to study potential links between ownership and productivity change. No definite indication of correlation is found. Veiderpass and Salas (1996) make efficiency and productivity comparisons between the cement producers in Peru and Colombia. They use a four input one output specification, and a chained input based Malmquist index for the period 1980-1986. The authors find that most production is characterized by CRS, and that productive efficiency as well as productive growth is generally higher in Peru than in Colombia.

3.3.5 The Electric Utility Industry

The productivity of electric utilities has been measured both with respect to production and distribution of electricity. Hjalmarsson and Veiderpass (1992), and Førsund and Kittelsen (1994) study the distribution of electricity. Both papers use the fixed based Malmquist index discussed in Section 3.2.6. Hjalmarsson and Veiderpass use data from 289 Swedish retail distributors during the period 1970-1986. They construct three different models which differ in the specification of outputs used, which allows them to distinguish among the buyers of electricity. Their main conclusions are:

i) In terms of electricity supplied, the productivity growth in

Swedish electricity distribution has been quite substantial during an extended period of time.

ii) In terms of number of customers served, the productivity growth has been negative.

iii) Ownership or economic organization does not seem to be related to productivity change in any significant way.

iv) The productivity growth in rural areas has been higher than the growth in urban areas, a fact probably due to the structural rationalization of rural electricity distribution, which led to many small service areas being merged into larger ones.

Førsund and Kittelsen (1994) study the Norwegian electricity distribution system and use data from two years, 1983 and 1989. Their study "shows an overall positive development of the magnitude of two per cent per year. Decomposing the total productivity development identifies frontier production shift as the main driving force." (p. 16)

Like in the above two papers, Førsund and Hjalmarsson (1995) use a fixed based Malmquist index and a nonparametric DEA approach for computing productivity. They study electric generating companies located in Southeast Asia. They find that the rate of productivity change over time (1981-1981) varies but that for most years there is technical progress.

In three different papers, Yaisawarng et al. study productivity of electric power plants in the U.S. All three papers use the nonparametric DEA approach in calculating the productivity scores. The first paper by Färe, Grosskopf, Yaisawarng, Li and Wang (1990) uses the Malmquist model (3.2.17) to study coal fired Illinois electric utilities for the period 1975-1981. They find no productivity growth on average. Klein, Schmidt and Yaisawarng (1992) using data from 54 mid-west coal fired electric utilities find that productivity: "fell at about a one percent annual rate from 1975 to 1979, and grew at about the same rate thereafter." (p. 228) (that is to 1987). Klein et al. using a fixed based model also studied the decomposition of productivity. They found that technical change fell with nearly two percent per year in the first period then gained by 1.24% thereafter.

Yaisawarng and Klein (1994) extend the above study by Klein et al. (1992) by using a subvector efficiency measure,[45] which allows for non-optimal scale and explicitly accounts for undesirable inputs and outputs. The data consists of 61 coal fired electric generating plants for the years 1985 to 1989. They find: "Comparing the years 1986 to 1989 to our fixed base year of 1985, the cumulative Malmquist index shows productivity retardation in 1986 and 1987. The first sign of recovery appears in 1988 and we see productivity growth in 1987." (p. 457.)

3.3.6 Transportation

The airlines and motor carriers are the two transportation industries which have been studied using the Malmquist productivity index. Starr McMullen and Okuyama (1996) use the nonparametric method to compute the Malmquist index (3.2.17) for U.S. motor carriers between 1976 and 1990. Since the nonparametric approach is nonstatistical, the authors use a bootstrapping technique to generate confidence intervals. They find significant technological decreases occurring in 1976-1978, 1979-1981, 1982-1985 and 1987-1989.

In three different papers, Sickles et al. study the airline industries in U.S. and Europe. Good and Sickles (1995) use both a stochastic semiparametric and a DEA approach to calculate distance functions for Eastern and Western European airline carriers. Their Malmquist productivity results point to a widening divergence in

[45] By subvector efficiency it is here understood that only a subvector of inputs are scaled, i.e., the input distance function is not defined on all inputs but just some.

efficiencies rather than productivities. Alam and Sickles (1995:a) analyse a panel of 11 U.S. airlines observed quarterly during 1970-1990 for cointegration and convergence. For a majority of firms pairs, cointegration can not be rejected. The authors are most interested in the efficiency change component of the Malmquist index (2.2.6), since it indicates how similar the carriers are becoming over time. They find evidence of convergence since the airlines on average, are close to the frontier. Alam and Sickles (1995:b) study the link between relative technical efficiency and stock market process with the same airline data as in (1995:a).

Distexhe and Perelman (1995) use a Malmquist model with attributes like in Färe, Grosskopf and Ross (1995:a) to study productivity among international airlines, including airlines from Europe, North America, Asia and Oceania (1987-1988). They find that only those airlines operating on a worldwide scale have been able to take advantage of technological progress.

3.3.7 Productivity in the Insurance Sector

Donni and Fecher (1995) provide an international comparison of productivity in the insurance sector for 15 OECD countries over the 1983-91 period. They use the geometric mean form of the Malmquist output based productivity index (3.2.17) and decompose productivity change into technical change and efficiency change. The authors impose variable returns to scale. Their model includes two outputs: life and non-life net premiums and one input: labor. They find considerable variation in annual technical efficiency as well as productivity. On average there was productivity growth, which was attributable to both components of productivity. They also investigate the determinants of productivity growth using a second stage probit model which includes as independent variables: ratio of reinsurance accepted, market share in OECD countries, life share of total insurance, foreign companies' market share, average company size and a time trend. Reinsurance and market share were positively related to technical efficiency, whereas size, life share and foreign market share were negatively related to technical efficiency.

In a country study, Fukuyama (1995:b) computes productivity of Japanese life insurance companies over the 1988-93 bubble-bust period. He uses the geometric mean Malmquist input-based index from (3.2.17) decomposed into scale, congestion and technical efficiency and technical change components. His outputs are reserves and loans, with inputs specified as peronnel, physical capital, and tied agents. Of particular interest was the effect of the bubble and bust on productivity in this industry and whether organizational form (stock vs. mutual) affected productivity. Based on nonparametric tests, the author concludes that organizational form did not matter, and that productivity improved over this time period, mainly due to technical change. Fukuyama (1995:c) estimates the Malmquist input-based index using a 1991-92 panel data of Japanese nonlife insurance companies. His finding shows that small nonlife insurers perform better than large or medium-sized insurers.

In another country study, Cummins, Turchetti and Weiss (1995) compute input-based Malmquist productivity for a balanced panel of 94 insurance companies in Italy over the 1985-93 period. They use the geometric mean formulation and disaggregate into technical change and efficiency change. One of the motivations for their study is the increasing competitive pressure in this industry due to membership in the EU, including the advent of 'bankassurance.' The authors include a careful discussion of the difficult issue of specifying a production model for this industry. They argue that in a general value-added framework, insurers provide three principal services: risk-pooling and risk-bearing, real financial services related to insured losses, and financial intermediation. They proxy these services using incurred losses for four categories of nonlife insurance, and life insurance benefits for life insurance, and the total value of invested assets (to proxy intermediation). Inputs include two types of labor, fixed capital and equity capital. They find overall a decline in productivity, attributable to technical regress. To investigate the determinants of technical efficiency, they include a regression with variables to reflect ownership type, line of business, asset mix, reinsurance role, etc. They find mutuals to be more efficient than stock firms, contradicting the expense preference model. Other interesting results include the fact that specialized firms were more efficient than diversified, it is better to be a net reinsurer, and more complicated investments and insurance type result in lower efficiency.

Using the same specification of outputs and inputs,[46] Cummins, Weiss and Zi (1995) compute input-based Malmquist productivity and also provide a measure of cost efficiency for U.S. stock and mutual property-liability insurers over the 1981-1990 period. They decompose productivity into technical change and efficiency change; cost efficiency is decomposed into allocative and technical efficiency. Since they are particularly interested in identifying any differences in technology between stock and mutual insurers, they also include an interesting exercise: they evaluate technical efficiency in a pooled setting, separate setting and then an 'out of sample' model in which firms from one group are evaluated relative to the frontier of the other group (without themselves being included in the reference technology). Their tests suggest that stocks and mutuals have different frontiers, and that each is dominant in producing the output vectors in which that group specializes, i.e., the frontiers intersect. Their cost efficiency results suggest that the stock frontier dominates the mutual frontier. The productivity results suggest that productivity improved in both mutuals and stocks, with slighly greater cumulated increases in mutuals. In both cases, improvements were due to technical change. They also computed productivity based on the cost model and found much greater cumulated productivity for stocks than mutuals, which they attribute to expense preference behavior in mutuals.

[46] The inputs are slightly different: labor, materials, policy-holder supplied debt and financial equity capital.

3.4 CONCLUDING REMARKS

Our review of the burgeoning literature related to Malmquist productivity indexes suggests to us that after a long hiatus following its inception in 1982 by CCD, this index has developed rapidly in terms of both theoretical dimensions and empirical applications. Many of the theoretical issues discussed in our review arose naturally as researchers attempted to use the index as an empirical tool.

One of the most fundamental issues raised in this way is the question of the appropriate restrictions to impose on the underlying technology when computing the component distance functions. Since one of the advantages of the index and the computational techniques used to compute it is the fact that few restrictions are imposed a priori concerning optimization or functional form, the natural inclination is to use the least restrictive model available, which would typically be the variable returns to scale model. As it turns out, intuition isn't foolproof. Our discussion of the theoretical role of constant returns to scale leads us to conclude that anyone who wishes to measure total factor productivity using the Malmquist index should, in effect, use constant returns to scale in specifying the reference technology. Otherwise, the connection between the Malmquist index and our general notion of total factor productivity as measuring average product (the ratio of outputs to inputs) is probably lost.

This does not mean that empirical researchers must ignore scale effects and performance relative to the variable returns technology. As shown by Färe, Grosskopf, and Lovell (1994) and Färe, Grosskopf, Norris and Zhang (1994), the Malmquist index computed under constant returns can be decomposed into changes in scale efficiency, changes in variable returns technical efficiency and technical change (computed relative to constant returns technology). This keeps the connection to TFP, avoids many computational problems which arise when computing mixed period problems under variables returns, and provides useful information.

From our theoretical review we also know that formulation of the Malmquist index to satisfy the circular test or to be of a fixed base type that is independent of the base year chosen, comes at a cost in terms of the underlying technology. Both of these require that technological change be 'neutral,' which imposes fairly restrictive structure on technology in the multiple output, multiple input case.

Another theoretical issue which arose from empirical practice, is the relationship between the CCD type Malmquist index, and what Diewert calls the Hicks-Moorsteen productivity index. Both types of indexes are composed of distance functions. The former compares observed data from two different time periods to a common frontier or technology (in the simple single ratio case). The latter is conceptualized as the ratio of an output quantity index to an input quantity index. As such, it is conceptually much closer to the Törnqvist and Fisher indexes than to the Malmquist index. In answering the question: when are the Malmquist and Hicks-Moorsteen indexes equivalent, we find that technology must be inverse homothetic (which implies a type of separability between inputs and outputs) and satisfy constant returns to scale. These are not trivial assumptions, and, given the close relationship between the Hicks-Moorsteen and Törnqvist/Fisher indexes, provides some insight into the restrictiveness of the conditions used by CCD to show equivalence between the Törnqvist and Malmquist indexes.

Thus we find that the Malmquist index is conceptually very general, including the Hicks-Moorsteen, Fisher and Törnqvist indexes as special cases. Because it does not require information on prices or shares, the Malmquist index does not require the optimizing behavior that would make those prices meaningful as aggregators. On the other hand, the Malmquist index does not naturally disaggregate into output and input quantity indexes, which may be of considerable policy interest. Both the Hicks-Moorsteen and Malmquist indexes require computation of distance functions, unlike the Fisher and Törnqvist indexes which do not require estimation of technology. This computational ease of these latter indexes is a very big advantage from the viewpoint of statistical agencies which must deliver their results on time, every time.

Turning more explicitly to the empirical literature related to Malmquist indexes, our sample of studies suggests that this has turned out to be an extremely useful analytical tool. The range of applications is wide: from individual firm data to country-wide international comparisons of productivity. Although we hope that this survey has provided some guidance for applied work (for example, with respect to the choice of a reference technology), researchers must still struggle with the following: what orientation should be chosen (input or output-based, for example), how should the model be specified in terms of inputs and outputs, what type of decomposition (if any) should be used, how should the results be reported, and how should the results be interpreted?

Some fairly general, practical advice can be gleaned from our survey. First, the orientation of the distance functions could be based on the presumed goal or benchmark for the decision making unit. For example, Swedish pharmacies are required by law to provide drugs at minimum cost. As a consequence, an input saving orientation, consistent with cost minimization was chosen. The choice of variables can be exceptionally challenging, such as in the case of insurance companies. As economists, we would like to see some connection to the underlying notion of outputs being related to possible revenue or benefits, and inputs tied to costs or resource use, and that the variables be exhaustive in some sense.

Decompositions should be employed when they fit in with the research question. By the same token, results should be presented in a way that relates to the research question. We found that studies which employed graphics in the presentation of results were more accessable and could be used to provide information on the distribution of results rather than just the first or second moments.

Other innovations we noted were efforts to test whether productivity changes were significant or significantly different across various groups using bootstrapping to construct confidence intervals, for example. Many studies also employed second stage statistical analysis to analyze their results and test various hypotheses. We expect to see refinements and extensions of these types of analysis in the future.

This survey has not addressed a number of issues that are potentially of considerable to those who work in this area. We have not included a discussion of Malmquist indexes based on representations of technology other than input and output distance functions. Possibilities include Malmquist indexes based on indirect distance functions (which are fortunately given careful attention in Bert Balk's contribution to this volume), as well as revenue or cost functions. There is also an emerging literature on specifying productivity indexes using what have come to be called directional distance functions. We have not included a discussion of the relationship of Malmquist productivity indexes to profit or profitability, nor adequately addressed the serious theoretical and empirical issues associated with quality. We realize that our survey will have missed many valuable contributions to this literature that we hope will be redressed in the next survey.

References

Aczél, J. (1990) "Determining Merged Relative Scores," *Journal of Mathematical Analysis and Applications* 150, 20-40.

Aigner, D. and S.F. Chu (1968) "On Estimating the Industry Production Function," *American Economic Review* 58, 826-839.

Alam, I. and R. Sickles (1995:a) "Long Run Properties of Technical Efficiency in the U.S. Airline Industry." mimeo, Rice University, Houston, TX.

Alam, I. and R. Sickles (1995:b) "The Relationship Between Stock Market Returns and Technical Efficiency Innovations: Evidence from the U.S. Airline Industry," mimeo, Rice University, Houston, TX.

Althin, R. (1995) *Essays on the Measurement of Producer Performance*, Ph.D. Dissertation, Lund University.

Atkinson, S. and P. Wilson (1995) "Comparing Mean Efficiency and Productivity Scores from Small Samples: A Bootstrap Methodology," *Journal of Productivity Analysis* 6, 137-152.

Balk, B. (1993) "Malmquist Productivity Indexes and Fisher Ideal Indexes: Comment," *Economic Journal* 103, 680-682.

Balk, B. and R. Althin (1996) "A New, Transitive Productivity Index," *Journal of Productivity Analysis* 7, 19-27.

Banker, R.D. and R.M. Thrall (1992) "Estimation of Returns to Scale Using Data Envelopment Analysis," *European Journal of Operational Research* 62, 74-84.

Bauer, P.W., A.N. Berger, G.D. Ferrier and D. Humphrey (1995) "An Empirical Investigation of Productivity Change in U.S. Banking: 1977-1988," University of Arkansas's College of Business Administration Working Paper No. 95-902.

Berg, S., F.R. Førsund, L. Hjalmarsson, and M. Suominen (1993) "Banking Efficiency in the Nordic Countries," *Journal of Banking and Finance* 17, 371-388.

Berg, S., F.R. Førsund, and E.S. Jansen (1992) "Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking, 1980-89," *Scandinavian Journal of Economics* 94, Supplement, 211-228.

Bjurek, H. (1994) "Essays on Efficiency and Productivity Change with Application to Public Service Production," *Ekonomiska Studier*, Göteborgs Universitet, Nr. 52.

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

Bjurek, H., U. Kjulin and B. Gustafsson (1992) "Efficiency, Productivity and Determinants of Inefficiency at Public Day Care Centers in Sweden," *Scandinavian Journal of Economics* 94, Supplement, 173-187.

Bjurek, H. and L. Hjalmarsson (1995) "Productivity in Multiple Output Public Service: A Quadratic Frontier Function and Malmquist Index Approach," *Journal of Public Economics* 56, 447-460.

Boisso, D., S. Grosskopf and K. Hayes (1996) "Regional Productivity and Efficiency in the U.S.: Effects of Business Cycles and Public Capital," Federal Research Bank of Dallas Research Department Working Paper: 96-02.

Burgess, J., Jr. and P. Wilson (1995) "Decomposing Hospital Productivity Changes, 1985-1988: A Nonparametric Malmquist Approach," *Journal of Productivity Analysis* 6, 343-363.

Byrnes, P., R. Färe and S. Grosskopf (1984) "Measuring Productive Efficiency: An Application to Illinois Strip Mines," *Management Science* 30, 671-681.

Cabezas Vega, L. and A. Veiderpass (1992) "Eficiencia Relative y Desarrollo de la Productividad en la Produccion Peruana de Cemento," *Economia*, XV, 195-227.

Cabezas Vega, L. and A. Veiderpass (1994) "Eficiencia y Cambio de la Productividad en la Industria Cementera del Perú. Aplicación de un Méodo no Paramétrico," *Trimestre Económico*, LXI, 309-333.

Caves, D., L. 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. and R. Färe (1994) "Hicks Neutrality and Trade Biased Growth: A Taxonomy," *Journal of Economic Theory* 64, 554-567.

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

Cummins, J.D., T. Guiseppe and M.A. Weiss (1995) "Productivity and Technical Efficiency in the Italian Insurance Industry," mimeo, paper presented at the 22nd Seminar of the European Group of Risk and Insurance Economists.

Cummins, J.D., M.A. Weiss and H. Zi (1995) "Organizational Form and Efficiency: An Analysis of Stock and Mutual Property Liability Insurers'," mimeo, Wharton School, University of Pennsylvania.

Devaney, M. and W. Weber (1995) "Rural Bank Efficiency and Contestable Markets," USDA Grant No. 9403254.

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

Diewert, W.E. (1992:a) "The Measurement of Productivity," *Bulletin of Economic Research* 44, 163-198.

Diewert, W.E. (1993) "The Measurement of Productivity: A Survey," Swan Consultants, Canberra.

Distexhe, V. and S. Perelman (1994) "Technical Efficiency and Productivity Growth in an Era of Deregulation: The Case of Airlines," *Swiss Journal of Economics and Statistics* 130, 669-689.

Domazlicky, B.R. and W.L. Weber (1997) "Total Factor Productivity in the States, 1977-86," *Journal of Regional Science* (forthcoming).

Donni, O. and F. Fecher (1995) "Efficiency and Productivity in the OECD Insurance Industries," CREPP discussion paper 95/03, Université de Liége.

Färe, R. E. Grifell-Tatjé, S. Grosskopf and C.A.K. Lovell (1997) "Based Technical Change and the Malmquist Productivity Index," *Scandinavian Journal of Economics* 99, 119-127.

Färe, R. and S. Grosskopf (1992) "Malmquist Productivity Indexes and Fisher Ideal Indexes," *Economic Journal* 102, 158-160.

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

Färe, R., S. Grosskopf and W.-F. Lee (1995) "Productvity in Taiwanese Manufacturing Industries," *Applied Economics* 27, 259-265.

Färe, R., S. Grosskopf and W.-F. Lee (1996) "Productivity and Technical Change: The Case of Taiwan," mimeo, Department of Economics, Southern Illinois University, Carbondale, IL.

Färe, R., S. Grosskopf, B. Lindgren and J.-P. Poullier (1997) "Productivity Growth in Health Care Delivery," *Medical Care* 35, 354-366.

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

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1989, 1994) "Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach," in A. Charnes, W.W. Cooper, A. Lewin and L. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston: Kluwer Academic Publishers.

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 (1994) "Produktivitetsutveckling inom Sjukhussektorn 1970-1992," *Den Offentliga Sektorns Produtivitesutveckling 1980-1992 Ds* 1994:17.

Färe, R., S. Grosskopf and P. Roos (1995) "The Malmquist Total Factor Productivity Index: Some Remarks," Center for Economic Studies, University of Munich, Working Paper No. 94.

Färe, R., S. Grosskopf and P. Roos (1995:a) "Productivity and Quality Changes in Swedish Pharmacies," *International Journal of Production Economics* 39, 137-144.

Färe, R., S. Grosskopf and P. Roos (1996) "Integrating Consumers Satisfaction into Productivity Indexes," Department of Economics, Southern Illinois University, Carbondale, IL.

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

Färe, R., S. Grosskopf, S. Yaisawarng, S.-K. Li and Z. Wang (1990) "Productivity Growth in Illinois Electric Utilities," *Resources and Energy* 12, 383-398.

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

Ferrantino, M.J. and G.D. Ferrier (1996) "Best-Practice Technology in an Adverse Environment: The Case of Indian Sugar, 1981-1986," *Journal of Economic Behavior and Organization*, 29:3, May 1996, 493-522.

Ferrier, G.D., M. Klinedinst and C.B. Linvill (1996) "Static and Dynamic Productivity Among Yugoslav Enterprises: Components and Correlates," mimeo, Department of Economics, 402 BADM, University of Arkansas, Fayetteville, AR 72701-1201.

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

Førsund, F. (1990) "The Malmquist Productivity Index," Department of Economics, University of Oslo, Memorandum No. 28.

Førsund, F. (1993) "Productivity Growth in Norwegian Ferries," in H.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency: Theory and Applications*, Oxford: Oxford University Press.

Førsund, F. (1996) "Productivity of Norwegian Establishments: A Malmquist Index Approach," in *Sources of Productivity Growth* Mayes D.G. (ed.), Cambridge: Cambridge University Press.

Førsund, F. and L. Hjalmarsson (1987) *Analysis of Industrial Structure: A Putty Clay Approach*, Stockholm, Almquist & Wicksell International.

Førsund, F. and L. Hjalmarsson (1979) "Generalized Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants," *Economic Journal* 89, 274-315.

Førsund, F. and S. Kittelsen (1994) "Productivity Development of Norwegian Electricity Distribution Utilities," SNF No. 10/94.

Frisch, R. (1936) "Annual Survey of General Economic Theory: The Problem of Index Numbers," *Econometrica* 3, 1-38.

Fukuyama, H. (1995:a) "Measuring Efficiency and Productivity

Growth in Japanese Banking: A Nonparametric Frontier Approach," *Applied Financial Economics* 5, 95-107.

Fukuyama, H. (1995:b) (forthcoming) "Investigating Productive Efficiency and Productivity Changes of Japanese Life Insurance Companies," *Pacific-Basin Finance Journal.*

Fukuyama, H. (1995:c) "Productivity Change and Technological Change in Depression: Evidence from the Japanese Non-Life Insurance Industry," *Fukuoka University Review of Commercial Sciences* 39:3/4, 893-912.

Good, D. and R. Sickles (1995) "East Meets West: A Comparison of the Productive Performance of East European and Western European Air carriers," mimeo, Rice University, Houston, TX.

Gouyette, C. and S. Perelman (1995) "Productivity Convergence in OECD Service Industries," CREPP working paper 95/04, University of Liège, Belgium.

Grifell-Tatjé, E. and C.A.K. Lovell (1994) "A Generalized Malmquist Productivity Index," Department of Economics, University of Georgia, Working Paper No. 94-410.

Grifell-Tatjé, E. and C.A.K. Lovell (1994) "Scale Economies and the Measurement of Producttivity Change: An Application to Spanish Savings Banks," *Perspectives del Sistema Financiero* 48, 123-132 (in Spanish).

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 (1995:a) "A DEA-Based Analysis of Productivity Change and Intertemporal Performance," *Annals of Operations Research*, forthcoming.

Grifell-Tatjé, E. and C.A.K. Lovell (1996) "Deregulation and Productivity Decline: The Case of Spanish Savings Banks," *European Economic Review*, 1281-1303.

Grifell-Tatjé, E. and C.A.K. Lovell (1997) "The Sources of Productivity Change in Spanish Banking," *European Journal of Operational Research* 98, 364-380.

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

Hjalmarsson, L. and F. Førsund (1995) "Productivity in Electricity Generation in Developing Countries: An International Comparison of Far Eastern Power Production." Paper presented for the International Conference on Energy and Environment, 8-10 May, Shanghai, China.

Hjalmarsson, L. and A. Veiderpass (1992) "Productivity in Swedish Electricity Retail Distribution," *Scandinavian Economic Journal* 94, S193-205.

Hjalmarsson, L. (1991) *Metoder i Forskning om Produktivitet och Effecktivitet med Tillämpningar på Offentlig Sektor*, Finansdepartementet Ds 1991:20.

Karlin, S. (1959) *Mathematical Methods and Theory in Games, Programming and Economics*, Reading: Addison-Wesley.

Kittelsen, S. and F. Førsund (1992) "Efficiency Analysis of Norwegian District Courts," *Journal of Productivity Analysis* 3, 277-306.

Klein, J.D., S. Schmidt and S. Yaisawarng (1992) "Productivity in the U.S. Electric Power Industry," in D.B. Audretsch and J.J. Sigfried (eds.), *Empirical Studies in Industrial Organization: Essays in Honor of Leonard W. Weiss*, Boston: Kluwer Academic Publishers.

Krug, E. (1976) *Stochastic Production Correspondences*, Meisenheim am Glan: Verlag Anton Hain.

Magnussen, J. (1994) *Hospital Efficiency in Norway: A Nonparametric Analysis*, Dissertation in Economics, No. 6, University of Bergen.

Mahler, K. (1939) "Ein Übertragungsprinzip für Konvexe Körper," *Casopis pro Pěstovăni Matematiky a Fysiky* 64, 93-102.

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

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

Nishimizu, M. and J.M. Page, Jr. (1982) "Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, 1965-78," *Economic Journal* 92, 920-936.

Ochoro, W. and L. Hjalmarsson (1995) "Productive Efficiency and Productivity Change in Kenyan Manufacturing," Chapter 2.

Odeck, J. (1993) "Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector, *Ekonomiska Studier*, Göteborgs Universitet, Nr. 44.

Pastor, J.M., F. Pérez, and J. Qusada (1994) "Efficiency Analysis in Banking Firms: An International Comparison," mimeo, Departmento de Análisis Económico, Universitat de Valéncia, Spain.

Perelman, S. (1995) "R&D, Technological Progress and Efficiency Change in Industrial Activities," *Review of Income and Wealth* 41, p. 349-366.

Piesse, J., C. Thirtle and J. van Zyle (1996) "Effects of the 1992 Drought on Productivity in the South African Homelands: An Application of the Malmquist Index," *Journal of Agricultural Economics* 47, 247-254.

Pittman, R. (1983) "Multilateral Productivity Comparisons with Undesirable Outputs," *Economic Journal* 93, 883-891,

Ray, S. and E. Desli (1995) "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries: A Deja Vu," Department of Economics, University of Connecticut.

Roos, P. (1993) "Three Essays on the Measurement of Productivity Changes," Licentiate dissartation, Lund University.

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

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

Starr McMullen, B. and K. Okuyama (1996) "Productivity Changes in the U.S. Motor Carrier Industry 1976-1990: A Malmquist Index Approach," mimeo, Department of Economics, Oregon State University, Corvallis, Oregon.

Studit, E.F. (1995) "Productivity Measurement in Industrial Operations," *European Journal of Operational Research* 95, 435-453.

Taşkin, F. and O. Zaim (1995) "Productivity Convergence, Technical Progress and Outward-Oriented Trade Regimes," Department of Economics working paper No. 95-8, Bilkent University, Ankara, Turkey.

Tauer, L.W. (1994) "Efficiency, Technology, and Productivity Changes on Individual Dairy Farms," Department of Agricultural, Resource, and Managerial Economics working paper 94-13, Cornell University.

Tauer, L.W. (1995) "Do New York Dairy Farmers Maximize Profits or Minimize Costs?" *American Journal of Agricultural Economics* 77, 421-429.

Tauer, L.W. (1996) "Productivity of New York Dairy Farms," mimeo, Department of Agricultural, Resource, and Managerial Economics, Cornell University.

Thirtle, C., D. Hadley and R. Townsend (1994) "Policy Induced Innovation in Sub-Saharan African Agriculture: A Multilateral Malmquist Productivity Index Approach," mimeo, Department of Agricultural Economics and Management, University of Reading.

Tulkens, H. (1986) "La Performance Productive d'un Service Public: Definitions, methodes de mésure, et application à la Régie des Postes de Belgique," *L'Actualité Economique, Revue D'Analyse Economique*, Montréal, 62, 306-335.

Tulkens, H. and A. Malnero (1996) "Nonparametric Approaches to the Assessment of the Relative Efficiency of Bank Branches," in D.G. Mayes (ed.), *Sources of Productivity Growth*, Cambridge: Cambridge University Press, 233-244.

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

Turk, J., J. Piesse and C. Thirtle (1996) "The Productivity of Private and Social Farms: Multilateral Malmquist Indices for Slovene Dairying Enterprises," *Journal of Productivity Analysis* 714, 447-460.

Vega, L.A.C. (1994) "Factor Substitution, Capacity Utilization, and Total Factor Productivity Growth in the Peruvian Manufacturing Industry," *Ekonomiska Studier*, Göteborgs Universitet, Nr. 53.

Veiderpass, A. (1993) "Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change," *Ekonomiska Studies*, Göteborgs Universitet, Nr. 43.

Veiderpass, A. and O. Salas (1996) "Relative Efficiency and Productivity Development in Colombian and Peruvian Cement Production," mimeo, Economics # 226, Göteborgs Universitet.

Weber, W. and M. Devaney (1996) "Bank Productivity Growth and Real Estate Exposure: The Credit Crunch Revisited," mimeo, Southeast Missouri State University, Cape Girardeau, MO.

Wheelock, D. and P. Wilson (1994) "Productivity Changes in US Banking: 1984-1993," Federal Reserve Bank of St. Louis, Research and Public Information Division Working Paper Series 94-021A.

Wheelock, D. and P. Wilson (1995) "Why are Banks Inefficient? Technical Progress and Productivity Change in US Banking, 1984-1993," mimeo.

Yaisawarng, S. and J.S. Klein (1994) "The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry," *Review of Economic and Statistics* LXXVI, 447-460.

4

IMPLEMENTING THE MALMQUIST PRODUCTIVITY INDEX: THE CASE OF THE NATIONAL CORPORATION OF SWEDISH PHARMACIES

by

Nils-Olov Norlander and Pontus Roos

4.1 INTRODUCTION

This essay looks at an ongoing attempt to implement Malmquist input-based productivity indexes for the National Corporation of Swedish Pharmacies (*Apoteksbolaget*, referred to below as "the Corporation"). The discussion proceeds from productivity calculations at the pharmacy level, distinguishing between productivity on the one hand and changes in productivity on the other. "Productivity" represents comparisons of the relative technical efficiency of pharmacies at a given point in time. The concept "change in productivity" stands for an alteration in productivity from one period of time to another; this alteration may be divided into sub-components such as changes in technical efficiency and technological change.

The process of implementation has been running for several years under a scheme of collaboration between the Chief Controller of the Corporation and the Institute for Health Economics (IHE). Two external researchers, Professors Rolf Färe and Shawna Grosskopf, are also involved in the project; they lend assistance over methodological problems and help us maintain high international quality in project work. We hope that the implementation process will come to be regarded as an example of successful cooperation between practitioners and the world of academic research.

In the early 1970s, retail trade in pharmaceutical preparations became regulated in Sweden. Under Swedish law, the Government decides who will be allowed to trade in druge. In consequence of these regulations, the National Corporation of Swedish Pharmacies was formed in 1971. From the outset, the Corporation has had a monopoly over the retail trade in drugs to the general public. The Government owns two thirds of the shares, the remainder belonging to the Corporation employees' pension fund. In 1995, the Corporation had a total sale of approximately SEK 20 billion and 12,000 employees.

The Corporation's monopoly over selling drugs to the public is regulated in contracts with the Government. Usually, these contracts run for five years, but there have been both shorter and longer periods. The present contract comprises two years. The contract with the Government outlines the overall aims of the relevant business activity, financial aspects included. The Corporation is obliged to conduct its affairs in such a way that a satisfactory supply of pharmaceuticals is provided at the lowest possible cost. The meaning of "a satisfactory supply of goods and services" is clarified in the contract. Pharmacies will procure and provide drugs prescribed by properly qualified physicians, dentists, veterinarians, or other authorized prescribers. In addition to drugs on prescription, pharmacies are required to provide pharmaceutical specialties and other useful drugs which may be sold over the counter. With respect to service, the relevant contract stipulates that the provision of drugs and services must cover the whole country and be adapted to local conditions. In addition, the contract exhorts the Corporation to disseminate information about pharmaceuticals to the general public, as well as to other agents in the health sector. When the

Corporation sells a drug or provides services, its prices must be such that the Corporation's costs are offset by income and reasonable interest is earned on the capital. In addition, when selling drugs to the public all pharmacies in Sweden must charge the same price. The monopoly over retailing to the public is the Corporation's predominant business area.

The contract mentions two other possible business areas for the Corporation. The first consists in supplying products to health-care producers. The terms governing these contracted services are settled in agreement between the Corporation and the producers concerned. The second amounts to sales of products other than pharmaceutical preparations and services naturally associated with the main activities of the Corporation. Under the contract, both these business areas are required to offset the costs they generate. There must be no subsidizing of any one business area by another.

Operations are conducted in over 900 pharmacies, approximately 800 of which conduct retail trade in drugs to the general public, so-called outpatient pharmacies. The main purpose of the remaining pharmacies, termed hospital pharmacies, is to supply pharmaceuticals and services to health-care producers. The Corporation's head office in Stockholm comprises, among other things, departments with an overall responsibility for finance, statistics, and the monitoring and planning of activities.

This study is restricted to discussing productivity and changes in productivity in outpatient pharmacies only. In a great majority of cases, drug retailing is the major business concern of these pharmacies, far beyond any other function. The extent to which health-care producers are supplied, on a contractor basis, varies from one pharmacy to the next. Sales of products other than drugs vary, too; certain pharmacies do a great deal of business along these lines and others hardly any at all. All business areas operate under the express requirement that customers be well served.

The present method of calculating productivity and changes in productivity in the individual pharmacy consists of two components: obtaining a quotient by dividing an index of transactions in the three business areas by the total amount of hours worked in the pharmacy; and examining changes in this quotient. In calculations of the index of transactions, the weight 1.0 has generally been employed, except in the case of transactions in the business area consisting in sales of other products than drugs; here the weight 0.4 has been used instead. In the ensuing discussion, this approach is referred to as "the Corporation method of calculating productivity and changes in productivity." The services produced by pharmacies are referred to as "outputs" and the resources available to pharmacies as "inputs".

The present method used by the Corporation of calculating productivity has been criticized, both internally and by external critics; the latter have mostly been researchers dealing with methods for measuring productivity. Intra-Corporation criticism has usually focused on the method's inability to capture *all* the services produced and provided by pharmacies, as well as on the fact that only the input of labor is taken into account. Furthermore, the weighting employed in calculations has come under attack, and some people have been unhappy about the exclusion of quality aspects. Among the practical problems brought up by critics are difficulties in achieving consensus about what weights to use, and the circumstance that pharmacies frequently feel that the method is unreliable at the pharmacy level.

External critics have disapproved of the index design used in the method, maintaining that the approximations it contains are too rough to be acceptable, not least when calculations refer to individual pharmacies in which the mixes of outputs and/or inputs differ. The problems involved in employing fixed weights over time, and in making comparisons across pharmacies, have also been noted. In addition, those problems over incentives which are inherent in the Corporation method have been criticized, and so has its imprecise associations with quality, costs, and profitability.

Methods based on Malmquist index approaches to the calculation of productivity and changes in productivity possess a number of essential advantages in comparison with the Corporation method. These advantages, and the criticism leveled at the present method, formed the point of departure in our attempt to implement new, Malmquist-based methods in the Corporation. Any attempt along such lines would, of course, have been impossible without the support of the Corporation management. We would like to express our great appreciation of the helpfulness shown by the people in charge of the Corporation, who not only supported an attempt at implementation for several years but also encouraged other projects closely related to our own.

4.2 IMPLEMENTATION OF MALMQUIST PRODUCTIVITY INDEXES

This section presents a historical account of the implementation process. The process comprises several steps which may be said to constitute a joint characterization of the implementation as such. Method development and the publication of methods in scientific journals, an important factor in this context, take place alongside the drafting and presentation of internal written and oral reports. It turns out that an understanding of methods geared to calculating productivity and changes in productivity, and to assessing links with costs, quality, and profitability, is to a great extent bound up with the competence and dedication of practitioners as well as with the "pride" in developing front-line control and monitoring systems for service-producing companies. The following pages describe, and comment on, the historical development of those studies and internal reports which make up the implementation process. A summary of these development is presented in Table 4.1.

When the Corporation management entrusted IHE with a commission in 1989, their action initiated the process of implement-

Table 4.1 A historical outline of studies all of which have played significant parts in the process of implementing Malmquist-based index approaches for the National Corporation of Swedish Pharmacies.

1989.	Study aimed at developing a new index for the measuring of productivity, and change in productivity, in pharmacies. Malmquist input-based indexes. Method and application published in Journal of Productivity Analysis; Färe, Grosskopf, Lindgren, and Roos (1992).
1991.	Internal project intended to compare Malmquist productivity indexes with the Corporation's own method of calculating productivity. Internal reports to the Corporation management.
1991.	Study geared to showing how Malmquist indexes can be expanded so as to include certain quality aspects of services supplied by pharmacies. Internal reports as well as publication in Journal of Production Economics; Färe, Grosskopf, and Roos (1995).
1992.	Initiation of pilot project for calculating average changes in productivity within the Danish pharmacy sector by means of Malmquist input-based productivity indexes. Internal report to the Danish Assocation of Pharmacists. Today, calculations based on Malmquist productivity indexes are being made on an annual basis.
1992.	Two internal projects in connection with a review of Corporation finances and activities. An initial project regarding proposals for future financial guidelines, as well as a second project concerning the supply of services in pharmacies.
1993-1994.	Project initiated with a view to testing the routine application of Malmquist index approaches when calculating productivity and changes in productivity in all outpatient pharmacies. Internal reporting to the management and controller unit of the Corporation.

Table 4.1 (continued) A historical outline of studies all of which have played significant parts in the process of implementing Malmquist-based index approaches in the National Corporation of Swedish Pharmacies.

1994.	Study aimed at showing how Malmquist input-based productivity indexes may be extended so as to include customer satisfaction with services provided by pharmacies. Internal report and scientific publication Färe, Grosskopf, and Roos (1996:a).
1994.	Internal report to the Corporation management on financial guidelines of importance to the Corporation's successful development. The report makes special mention of a logically constructed system for the measuring of productivity, quality, and profitability.
1994.	Study aimed at elucidating the possibilities of assessing productivity for a group of pharmacies operating an internal network of services. Internal report as well as publication of methods and application; Färe, Grosskopf, and Roos (1996:b).
1995.	Current internal reporting on pharmacy productivity to the Corporation's Chief Controller.
1996.	Study intended to demonstrate the linkage between Malmquist productivity indexes, costs, and profitability. Internal reports and publication of method and application; Färe, Grosskopf, and Roos (1996:c).

ing Malmquist productivity indexes in the Corporation. What the Corporation wanted IHE to find out was whether new pharmacies were more productive than older ones, looking primarily at effects on pharmacy productivity. After an introductory meeting, it was apparent that the Corporation method of measuring productivity was not an acceptable starting-point for analyzing productivity and modern pharmacy design. A productivity index hence had to be developed as an initial step in the project, and it resulted in a Malmquist input-based productivity index. The index method and its application to a fairly limited selection of pharmacies were published in *Journal of Productivity Analysis*; see Färe, Grosskopf, Lindgren and Roos (1992).

The Corporation viewed the development of a new index for measuring productivity with interest, but also with a degree of skepticism. Questions were asked, such as "How does the Malmquist index relate to our own index? Will results be dissimilar? Can both methods yield the same results? How are we to understand the principles behind the calculations?" Such queries resulted in a parallel project involving the Corporation and IHE. This new project was intended to compare the Malmquist index approach with the Corporation's method of gauging productivity, elucidating the advantages and drawbacks of each method. An internal report was presented to the Corporation management in the autumn of 1991. It is clear from this report that the two methods did tend to yield different results, partly because calculations would frequently be based on different variables; thus, for instance, the Corporation method restricts input to hours worked. The report presents outcomes for a selection of pharmacies. For some of them, the two methods yielded similar results; but for others results were quite dissimilar, favorable productivity development according to the Corporation method turning into an unfavorable trend when calculations were based on the Malmquist index. Even when calculations rested on exactly the same data, these differences would largely remain – a circumstance which naturally generated questions regarding differences in the make-up and characteristics of the methods concerned. The report also tries to explain why the particular characteristics of the two methods may result in their yielding very dissimilar outcomes. In addition, the report shows how a division may be made into efficient and inefficient pharmacies, and how the method can be used to create helpful points of reference for the latter. In addition, the report includes a brief outline, without empirical estimates, of ways in which such qualitative aspects as waiting times might be taken into account. The conclusions of this report emphasize the advantages involved in employing a Malmquist productivity index; for instance, it does not use fixed weights for the aggregation of inputs and outputs, does not call for uniform measuring units in variables, offers an opportunity to identify inefficient and efficient pharmacies, and makes it possible to divide productivity development into sub-components.

The issue of qualitative aspects and their connection with productivity was one that interested the Corporation management. The contract with the Government regarding the monopoly over retailing drugs comprises explicit requirements as regards the services offered by pharmacies. Naturally, these services entail expenditure for the pharmacies that supply them. A project aimed at showing how quality aspects can be taken into account with the Malmquist productivity index was initiated in 1991. The quality aspects to be considered were different measures of accessibility pertaining to goods and services.

The project developed what might be termed Malmquist input-based productivity and quality indexes. In the relevant article, data from a selection of pharmacies illustrate this index method. The study was later published in *International Journal of Production Economics*; see Färe, Grosskopf, and Roos (1995). The selection of pharmacies – a third of all the existing ones – was not undertaken at random. It was restricted to those pharmacies which had regularly reported applicable data regarding waiting times, hours of business, and the proportion of all submitted prescriptions to have been dealt with within a 24-hour period. These data are collected by the Corporation head office in connection with an annual inventory in which pharmacy operations are scrutinized. At the relevant point in time, there was no analysis that might have indicated a connection among these variables, referring to the services provided by a pharmacy, and that pharmacy's productivity, or any other performance indicator. There was, however, a clearly stated wish to demonstrate how qualitative aspects may be taken into account in a calculation of productivity, as these variables are associated with pharmacy activities which demand plenty of resources. The records turned out to contain seriously flawed data, though.

Sometimes entries were missing, and/or the information supplied was obsolete.

The Corporation management took a very favorable view of the development of a Malmquist productivity index. The opportunity to include all inputs had been known before; but now the Corporation found itself in a position to pay far greater attention to pharmacy output in accordance with the Government contract. Productivity calculations were able to catch alterations in the services provided by the pharmacies. Now, for instance, productivity calculations were able to show to what extent a change was due to alterations in quality. The Malmquist productivity index could be divided into three components: (1) a change in efficiency; (2) a change in technology; and (3) a change in quality. When quality aspects of this kind can be included in productivity calculations, we would also expect the foundations to have been laid for a control and monitoring system able to give thorough support to the purposes and policies of pharmacies.

In 1992, the Corporation undertook a review of its operations and their efficiency. Among all the reports that resulted from this inquiry, two – both submitted by IHE – should be mentioned in this context. The first draws up future economic guidelines intended to safeguard the long-term efficiency of the Corporation. Among the points made by this report is a change in the direction of more decentralized decision-making within the Corporation, and an intention to retain a uniform structure conducive to enhancing efficiency. In future, as local variations become increasingly important, we may expect that more will be demanded of the monitoring systems which the Corporation employs in order to measure pharmacy productivity. In this regard, the Malmquist index for measuring productivity possesses essential characteristics which should be taken into account when monitoring methods are chosen.

A second IHE study deals with the services provided by pharmacies. An internal report addressed to the Corporation makes it clear that both the number of services and their extent vary markedly from one pharmacy to the next. In certain pharmacies the supply of services is considerable, whereas others supply none or next to none. For many of these services, pharmacies have not set up payment routines; they perform them referring to interpretations of central contracts between the Government and the Corporation. Consequently, the services concerned are frequently financed by means of a mark-up on the price of goods sold inside the relevant business area. Of course, an inventory of pharmacy operations is essential when variables in productivity models are to be identified, which was facilitated by this report.

Thanks to the methodological developments mentioned above, to the reports about them, and to the encouraging way in which the Corporation management received them at this time, it became possible to launch a project aimed at creating an operationally usable model for monitoring pharmacy efficiency. Here it should be pointed out that the Corporation asked whether other organizations employed Malmquist-based index approaches. Our reply was that we did not know of any organization that had implemented these methods, but that a large number of studies existed and that the Corporation was very likely to come to be regarded as a pilot case. We were able to draw attention to an important exception. In Denmark at this time, a discussion was going on between the Danish Association of Pharmacists, the National Danish Board of Health, and the Danish Treasury concerning modern methods for the measuring of productivity changes in the Danish pharmacy sector. In the course of these conversations mention was made of the method-development work going on in Sweden, and IHE was commissioned, as a pilot project, to calculate average changes in productivity in Danish pharmacies with the aid of a Malmquist input-based index.

In 1993, the Corporation and IHE jointly set up a project intended to monitor all outpatient pharmacies routinely, relying on a Malmquist productivity index. The purpose of this project was to calculate productivity and productivity development pertaining to all outpatient pharmacies – over 700 – for a period of two years. Calculations were to be made every year, as well as every six months on a rolling basis.

An essential first step was taken when variables for inputs, outputs, and quality attributes were defined and it was decided how these variables should be measured. The definition of variables was made in collaboration with pharmacy managers, the controller unit at the Corporation, and IHE. Taken altogether, 8 inputs, 7 outputs, and 3 quality attributes were defined. For most of the variables it was fairly easy to determine the appropriate measuring unit, for example hours worked by different categories of staff, purchased services in fixed prices, and other operative expenses, also in fixed prices.

The biggest problem arose over the ways in which pharmacies used capital. In order to come up with an applicable measurement for capital, it was necessary to calculate the capital stock value of pharmacy equipment, mainly shelving systems, counters, and other fixtures and fittings on the premises. The method employed for this purpose assumes that capital efficiency was the same on the first day as on the last on which pharmacies used the relevant object. Each pharmacy keeps a rolling list showing exactly which fixtures and fittings, and which other pieces of equipment, are actually in use. Furthermore, this list contains information on the year of purchase and on nominal costs. Nominal capital costs were deflated with the aid of a price index suitable for the purpose. The capital objects whose value, in fixed prices, was calculated in this way were then summarized to form a total measure of the pharmacy's capital. The flow of capital services is assumed to be directly proportional to the calculated capital stock.

The project also included the calculation of productivity and changes in productivity, with and without quality variables, and the pinpointing of differences across pharmacies. Some 10 per cent of pharmacies dropped out, mostly because of incomplete data on capital and quality variables. Even so, the drop-out rate was considerably lower than it had been in previous years. The results from these calculations were presented, in written and oral internal reports, to the Corporation management and controller unit. The reports contained detailed statements for each pharmacy as regards data used; efficiency; potential for inefficient pharmacies; and changes in productivity, both as a total and divided into components. The calculations were also compared to those results that had been obtained when the Corporation's own method was applied. Whenever there were considerable discrepancies, attempts were made to account for them. Thus, for example, one explanation might be that a certain pharmacy had incurred unreasonably high costs in connection with purchased services or excessive capital costs, that is to say, inputs which cannot be taken into consideration when using the Corporation method of productivity calculation. The results also showed that some pharmacies were at a disadvantage if evaluations of productivity failed to take all outputs and quality aspects into account.

In 1995, the calculations that had been performed in 1993-1994 were transformed into current reporting to the Corporation Financial Director (CFO) and Chief Controller. Besides containing the results mentioned above, these semiannual reports offer information on how many times a pharmacy has been involved in designing frames of reference for inefficient pharmacies. Out of a total of over 700 outpatient pharmacies, a fairly small number (20 - 30) have turned out to have been very frequently used as references. This information is of value for the controller unit from a "benchmarking" point of view. The calculations which include linear programming help to identify pharmacies that succeed better than others, as well as to determine what characterizes these pharmacies. Presentations of results and methods for individual pharmacies run parallel to the reports sent to the controller unit. Communication with pharmacies proved to be of major importance when it came to persuading pharmacies to accept the method, controlling quality in data reporting from pharmacies, and securing views on the variables employed. In addition, all middle managers – in this context, the managers of pharmacy groups – have obtained a sufficient amount of training for the method to be put to operative use.

While this full-scale test of the Malmquist index method is being wound up, further questions are raised with a bearing on the following issues: (1) Productivity and customer assessment of pharmacy activity; (2) the link between productivity, costs, and profitability; and (3) the economic consequences of co-operating in pharmacy groups. The first question has been analyzed in a research project, financed by the Corporation, aimed at investigating the possibility of bringing customer assessment of pharmacy services into Malmquist approaches to measuring productivity. A recently completed study, comprising method development and application to a minor selection of pharmacies, indicates an approach which makes it possible to consider local variations with regard to service as well as customer assessments of services provided; see Färe, Grosskopf, and Roos (1996:a). In this variant of a Malmquist productivity index, decomposition can be undertaken with reference to changes in technical efficiency, changes in technology, and changes in quality. Changes in quality include changes in usefulness to customers within the framework of the services provided by the pharmacy. It should be noted that the method pays attention to local variations in customer valuation of pharmacy service. One significant contribution made by this study consists in the method's showing how an assessment of a pharmacy's production may be made dependent on the valuations of customers without calling for observations in the form of market prices. This is an essential point, as there are quite a few cases in which no market prices exist and others where such prices cannot be used for the purpose.

The linkage between productivity, quality, and profitability in the case of regulated operations, such as the activities of the Corporation, is being studied in an ongoing project financed by the Corporation. The issue was raised by the Corporation's Financial Director (CFO) in the winter of 1995/96. The method proposed by IHE makes it possible to analyze whether an observed change in profitability may be related to changes in productivity and quality; see Färe, Grosskopf, and Roos (1996:c). The importance of a monitoring system able to perform these functions was previously brought out in an internal report to the Corporation's Chief Controller and put before the management. This report emphasizes the virtues of a monitoring system which does not lead to conflicts between costs, quality, and profitability, and the value of a system which engenders incentives to pharmacies to develop in the desired direction.

Work is in progress on the development of a model for analyzing the value of cooperation in pharmacy groups. This study was initiated by the Operations Manager of the Corporation. Pharmacies are organized in groups comprising some 20 to 30 pharmacies each. There is also an exchange of services between pharmacies. The question which the study is intended to address is whether this local network generates a productivity for the group as a whole which is greater than the sum of the individual pharmacies' productivity; see Färe, Grosskopf, and Roos (1996:b).

This historical description of scientific works and internal reports was intended to describe a number of events which have played a part in implementing Malmquist index approaches in an organization like the Corporation. In fact, the connection between theoretical models and operational instruments for the organization has been, and remains, an essential prerequisite for successful implementation. Another and equally important prerequisite is that the organization itself is able to perceive and identify the need for change. Many of the publications and reports mentioned above evolved out of finished projects as attendant questions; this suggests that the need for change may initially be hard, or even impossible, to perceive but that it becomes increasingly evident as the process continues. Consequently, the dissemination of knowledge is part and parcel of the implementation itself. One practical example is found in the training scheme for Corporation staff which is taking place alongside method development and calculations based on the methods evolved.

4.3 PERFORMANCE INDICATORS, INCENTIVES, AND MALMQUIST INDEXES

The financial department at the Corporation is keen to ensure that the performance indicators used to monitor and control operations supply incentives for the Corporation's development in useful directions. This might seem too self-evident to mention; but in fact it is not always a straightforward matter. The ensuing section looks at ways in which the application of Malmquist productivity indexes reduces the danger of conflicts between performance indicators in situations where each of them throws light on a certain aspect of operations, but they are not independent of one another. The current monitoring of pharmacy activities consists of performance indicators concerning (1) total costs; (2) profitability; and (3) the various services provided by pharmacies. The following discussion is restricted to these three performance indicators and their relations to productivity.

The total costs of the Corporation include labor, capital, external services, and other operational expenses. Total cost is determined by the pharmacy's production volume, prices of inputs, and pharmacy productivity. Profitability stands for the difference between income and expenditures – i.e., the profit taking account of certain restrictions, contained in the contract with the Government, with respect to what is felt to constitute a reasonable return on capital. The indicators of service that are relevant in this context have two things in common: all of them require resources, and they cannot be aligned with products sold to customers. In addition, they all constitute qualitative aspects of pharmacy service, which means that they possess some value in the eyes of customers. The weekly business hours of a pharmacy may be regarded as an indicator of the availability of goods and services. Extended hours call for more resources. Other measured indicators are the proportion of prescriptions distributed within 24 hours, queuing times for pharmacy customers, and the number of customers who received the wrong products. The latter category indicates the level of safety in pharmacy distribution to customers. For example, routines such as counter-signing can be expected to reduce the risk of the customer's failing to obtain the drugs prescribed. Safety systems demand resources, to, entailing consequences for costs, quality, and productivity.

Up to 1996, the Corporation management has taken a number of sequential steps aimed at improving the control of productivity and profitability and reducing the risk of conflict between performance indicators. Some of the steps have been tested in practice; others were intermediate stages of a rather more transitory character. Among the tested steps, the following measures might be mentioned:

a) targeting the number of prescriptions handled, or a weighted index of outputs, in relation to hours worked;

b) operating several performance indicators according to a so-called Balanced Scorecard;

c) trying to weight a number of performance indicators so as to obtain a singled "grade" on a predetermined scale (as in quality certification), and attempting to identify the interdependence of underlying variables;

d) testing a Malmquist productivity index;

e) making a Malmquist productivity index comprise aspects of quality as well; and

f) linking a Malmquist productivity index to profitability.

First, some comments should be made on this maturative process. The original approach, item a), which focuses on the number of prescriptions dealt with in a working hour, amounts to concentrating the largest mass of expenditure and on variables that are easy to understand. The drawback, however, is the difficulty of dealing with various inputs without transforming them into a uniform measuring unit. Likewise, it is hard or even impossible to weight outputs and quality aspects successfully; such a proceeding would also require a uniform measuring unit for variables, as well as consensus on what weights to use. In the matter of consensus regarding weights – for instance, weights determined by measuring time used for various activity componens – serious problems arise. These weights are bound to be strongly related to average conditions for all pharmacies, and consensus may be difficult to achieve as some pharmacies will, with reason, regard themselves as "losers". When it comes to coping with several outputs, several inputs, and quality, the method has so many drawbacks that it can hardly be said to yield more than a "half-educated guess". Item b), the Balanced Scorecard, started life as a complement to financial performance indicators. According to this variant, it is important to make measurements at an early point in time and not reactively, for instance on finding that operations have run at a loss. With the Balanced Scorecard, steering becomes more proactive. Its disadvantage is that a Balanced Scorecard is presented as a palette of performance indicators where "target values" may be hard to set independently of one another. The result will be a vague idea of what might constitute a good performance; in consequence, goals are less likely to be achieved. The notion of weighing different performance indicators in ways results in one universal performance indicator, as in item c), has been launched as a desirable approach. The fact that it provides a "total quality measurement", based on various certification programs, certainly constitutes a strength. In these programs different dimensions are graded separately, wholly independently of one another. A summation to an index figure ensues, with the aid of established weights. In practice, however, a several high-ranking business have soon found themselves with untenable relations between outputs and inputs. The trouble with established models of this kind is equally obvious: negotiated and/or fixed weights, without connections between inputs, costs, output, quality, and profitability, do not allow for any steering, because there is a risk that these models may reflect underlying success factors in a random manner. Such objections have much in common with the ones directed against a) above.

Another challenge emerged instead: that of constructing a descriptive model able to indicate links between inputs, outputs, quality, and profitability, as well as changes over time, in advance. Malmquist indexes according to items d) - f) form an attempt at designing such a descriptive model.

Disregarding quality aspects to begin with – or assuming their relations to pharmacy output to be fixed – we find an unambiguous connection between the observed cost of an individual pharmacy and its productivity. If, for instance, productivity indicates a 10 percent potential in improved productivity, this is synonymous with a potential cost reduction amounting to 10 percent, input prices being given. This simple link with costs arises in consequence of the duality inherent in Malmquist input-based indexes and the cost function of the pharmacy. If the above-mentioned quality aspects of pharmacy service are not always in a fixed relation to outputs – which would seem to be a reasonable contention – it is essential for productivity calculations to take this into account. A Malmquist productivity index can be extended in such a way that quality indicators form part of the calculations. Here, too, thanks to these indicators being included, a clear connection with costs emerges. A cost is to be regarded in relation to the achievement of a certain output and quality. A 10 percent productivity improvement is tantamount to a 10 percent cost reduction at fixed input prices. When the cost function is also made to contain qualitative aspects, the above-mentioned duality makes the linkage – known in advance – between the Malmquist index and the pharmacy's cost possible.

In what way, then, are productivity, quality, and profitability interconnected? An answer to that question calls for a description of the Corporation's pricing regulations on goods and services. With respect to the goods supplied by pharmacies, the price is very largely set by means of a mark-up on the cost price. For an individual item, the same price is charged all over the country. Services are either priced according to standard assessments of what it costs the business to produce and provide the relevant service, or by way of a mark-up on the cost of the product concerned. In the latter case, there is hence no direct compensation. It is reasonable to expect that the difference between a pharmacy's income and its corresponding costs, in other words its profitability, will be affected by imbalances between production costs and prices used. Thus the pricing regulations may per se account for differences in the observed profitability of pharmacies. Pricing strategies, and the fact that the three business areas operate different strategies, must be taken into account when the observed profitability of pharmacies is subjected to comparisons; otherwise there is a danger of "erroneous" information. Today, no model employed by the Corporation is able to demonstrate the linkage between profitability, costs, quality, and a Malmquist productivity index. It should be pointed out, though, that a certain development of a Malmquist index of profitability, costs, and quality was recently presented; see Färe, Grosskopf, and Roos (1996:c). Among other things, this report points to an approach which makes it possible to calculate the extent to which an observed change in observed profitability can be referred to changes in productivity and quality.

We wanted this section to demonstrate that Malmquist productivity indexes possess a vital quality in their natural connection with costs, quality, and profitability. The flexibility of these indexes has been indicated, too. For management purposes, a Malmquist index constitutes an instrument which may be said to point in the direction of a link between financial control and the steering of activities.

In comparison with the present performance indicators for productivity, costs, quality, and profitability, Malmquist productivity indexes represent a substantial improvement. Today, there is a patent risk that different messages will go out to pharmacies because performance indicators are calculated separately, despite the fact that they are strongly dependent on one another. One example may be quoted in order to illustrate these conflicts. The financial department regards a reduction in costs for a pharmacy as a favorable development. The trouble is that the cost was calculated without paying attention to all outputs and changes in quality. At the same time, the quality department reports a falling-off in the services supplied by the pharmacy, calling for improvements which the pharmacy manager interprets as demands for more resources and hence higher costs. Another department sends a report to the same pharmacy telling it that profitability has decreased and asking both for cost retrenchment and improved sales. Nowhere has the connection between productivity, quality, and pricing regulations been taken into account. This simple illustration will, we hope, indicate the practical problems that face a controller when the performance indicators he/she wishes to use are not synchronized. The danger of mismanagement, a steady loss of confidence within the pharmacies themselves, and unclear incentives are obvious in this example.

4.4 CONCLUDING REMARKS

Why does the Corporation's Chief Controller advocate the implementation of Malmquist-based indexes as a vital instrument of financial monitoring and control? The answer to that question rests on two factors: the implementation strategy, and the advantages of Malmquist indexes in relation to other methods.

From a management point of view, Malmquist index approaches have two essential qualities: Malmquist indexes have turned out to be very *difficult to manipulate*; and they provide *correct incentives* for productivity, costs, quality, and profitability. Resistance to manipulation is at least as important as the choice of a control system which possesses adequate logical characteristics and targets essentials. The best system, with the best intentions, may soon founder because of manipulation followed by incentive problems and conflicts. The dual properties of Malmquist-based approaches, which make it possible to show in advance how changes are likely to affect productivity, quality, and so on, also turned out to be instrumental in preventing manipulation and providing pharmacies with "rules of the game" from the outset. Being in possession of such rules means that the manager and staff of a pharmacy are able to identify how the management distinguishes between desirable and undesirable developments. With such a monitoring instrument, the link between finance and activity becomes evident. It is, for instance, clear how the management assesses an increase in quality when costs have gone up as well, or a productivity improvement when quality has been reduced at the same time.

Even when the need for an improved system of financial monitoring and control is widely felt and identified by many people in a business, it may often be hard, or even impossible, to implement Malmquist index approaches because employees fear being perceived as losers. With regard to this component of the implementation strategy, it was hence essential for the *right number of people* to be chosen for the purposes of achieving acceptance and securing dissemination and for selecting *the right type* of persons when it came to attaining consensus on the logic of control and steering. Among the things that characterized the persons concerned was the fact that they would ask questions on how methods could be extended and made to comprise increasing amounts of information which should, in their view, be included in or related to the analyses. An initial model based solely on inputs and outputs was followed by questions regarding aspects of quality and, in due course, profitability. The very flexibility of Malmquist index approaches, i.e., their inherent expandability, played an important part in persuading these people to support methods which did not initially take all aspects into account. With regard to data, there was no postulated requirement that the methods used in measuring productivity should be based on existing data only. To the relevant persons, the decisive thing was the realization that calculations would be feasible if data were available. The link with data ensured that a growing number of pharmacies could be analyzed as the productivity-measurement database was extended and data quality improved.

An implementation process may also be considered against the question of whether that process was impelled by internal and/or external pressure. In the case of the Corporation, implementation was not a directive issued by the management in order to meet external demands for greater good-will or profit. Instead, it was the Corporation's own organization, more particularly its controller unit, that called for better monitoring and control instruments, a wish supported by the management. Besides, the people concerned observed an opportunity to improve external accounts, therby anticipating possible future requirements concerning, for example, clearer information on the development of costs and productivity in pharmacies.

The factual effects of implementing Malmquist-based index approaches remain to be evaluated. Such an evaluation cannot be made until the index approach has been in use for a couple of years, the full consequences of the method being known to and accepted by pharmacies. This is the situation on 31 December 1996: a full-scale test has been performed, and the productivity-calculation methods are operatively ready for use by the controller unit. It will be easier to apply them more comprehensively at the pharmacy level once computer programs are available which enable the individual pharmacy to carry out calculations on its own. It seems most urgent to push ahead with a PC-based program for learning and performing the relevant calculation methods; such a program may well be a necessary prerequisite for ensuring that this type of financial analysis is accepted. Lack of practical experience with Malmquist index methods as an up-to-date instrument for financial monitoring remains the greatest obstacle to implementation, both in the Corporation and in other organizations. Hence it is a source of satisfaction to us that other organizations are also beginning to show an interest in the operational use of Malmquist index approaches. There are both private businesses and public authorities among these pioneers. The former perceive advantages in applying Malmquist index approaches as an internal instrument, but they also see it as a useful method in external accounting. Public bodies are more interested in the ability of this instrument to elucidate and control the development of private and public organizations, not least with regard to production and to the supply of services.

References

Färe, R., S. Grosskopf, B. Lindgren and P. Roos. (1992). "Productivity Changes in Swedish Pharmacies 1980-1989: A Nonparametric Malmquist Approach," *Journal of Productivity Analysis*, pp. 85-101. Also published in *International Applications of Productivity and Efficiency Analysis*, T. Gulledge and C.A.K. Lovell (eds.) Kluwer Academic Publishers (1992) pp. 81-97.

Färe, R., S. Grosskopf and P. Roos. (1995) "Productivity and Quality Changes in Swedish Pharmacies," *International Journal of Production Economics*, pp: 137-147, (1995). Also published in *Intertemporal Production Frontier: With Dynamic DEA* by R. Färe and S. Grosskopf, Kluwer Academic Publishers (1996).

Färe, R., S. Grosskopf and P. Roos. (1996:a). "Integrating Consumer Satisfaction into Productivity Indexes." IHE Working Paper 1996:4.

Färe, R., S. Grosskopf and P. Roos. (1996:b). "Network and Productivity Models of Swedish Pharmacies." Mimeo.

Färe, R., S. Grosskopf and P. Roos. (1996:c). "Profit Productivity and Quality: A Directional Distance Function Approach." IHE Working Paper, 1996:5. Also forthcoming in "Methodological Issues in Official Statistics", Statistics Sweden (SCB).

The National Corporation of Swedish Pharmacies (Apoteksbolaget AB), Internal Reports.

5

MALMQUIST PRODUCTIVITY INDEXES: AN EMPIRICAL COMPARISON

by

Hans Bjurek, Finn R. Førsund, and Lennart Hjalmarsson

5.1 INTRODUCTION

The interest in empirical applications of the Malmquist productivity index, proposed by Caves et al. (1982) (CCD), took off after Färe et al. (1994a) (originally circulated as a working paper in 1989) showed the relative ease with which the index could be calculated in the case of piecewise linear frontiers. The additional attraction added by Färe et al. (1994a) was to maintain inefficient operations (accommodated initially in the CCD definition, but then ruled out by assumption). Moreover, they decomposed productivity change into efficiency improvement and frontier shift in the general case of multiple outputs and inputs, generalizing the parametric single output approach in Nishimizu and Page (1982).

It is worth noting that when the production units are allowed to be inefficient, total factor productivity change is no longer equal to technical change measured by a shift in the technology, as is the case if units are efficient. Average performance catching up with best practice as a source of productivity improvement is an issue with a long tradition (see, e.g., Førsund and Hjalmarsson, (1987), chapter 1).

The general understanding of total factor productivity (TFP) is that it is measured by an index of outputs divided by an index of inputs, or as the shift in the production function (Diewert, 1981). These two approaches are identical when the production function is defined on continuous time and for efficient production units. The nature of returns to scale is not limited in these definitions. We may have variable as well as constant returns to scale production functions. A stated purpose of CCD, when defining the Malmquist Productivity Index for discrete data, was to introduce index number measures of productivity without having to approximate productivity concepts defined with respect to continuous time. The index measures require explicit knowledge of production technology, but no restriction was placed on returns to scale properties. However, consequences of working with variable returns to scale technologies were not elaborated upon.[1]

A few remarks on the relevance of variable returns to scale may be inorder. Indeed, Salter (1960), pointed to scale economies as a source of productivity improvement: "A striking example of the importance of such economies [of scale] is provided by the development of the grid system in electricity supply. ...There can be little doubt that such economies were one of the most important factors in the rapid rate of productivity increase in the electricity industry." (p. 141). This is exactly the sector we have studied here. There are other technologies also characterized by scale economies, to quote Denny et al. (1981): "For regulated industries, the assumptions of constant returns to scale and perfect competition are likely to be particularly inappropriate." (p. 180). Some industries, like distribution of electricity, are regulated just because they

[1] It should be mentioned that when the connection between a geometric mean of two Malmquist indexes and the corresponding Törnqvist index was established, a scale elasticity change term appears on the Törnqvist index side in the general case of variable returns to scale of the Malmquist indexes.

exhibit increasing returns to scale. On a broader scale, empirical evidence of variable returns to scale for a wide range of industries is provided in, for example, Haldi and Whitcomb (1967), Pratten (1971) and Caves et al. (1981). Thus, there is no doubt that a productivity index should be able to measure productivity changes even for such technologies.

Grifell-Tatjé and Lovell (1995) showed in a simple single input single output numerical example that the Malmquist productivity index does not correctly measure changes in TFP in the presence of variable returns to scale. To remedy this shortcoming Bjurek (1994 and 1996) suggested an alternative definition of the Malmquist productivity index as a ratio of a Malmquist output quantity- and a Malmquist input quantity index. The components of the quantity indexes are efficiency measures or distance functions. Diewert (1992, p. 240) briefly comments on the possibility of defining four different theoretical productivity indexes (combinations of two output- and input indexes respectively corresponding to the two period technologies) based on Malmquist quantity indexes without further developing this approach. Diewert noted that these indexes were geometrically defined by Moorsteen (1961), who discussed approximations of productivity given information on both prices and quantities in the two-input two-output case.[2]

Extending the definition in Moorsteen to allow for inefficiency is straightforward, but implies that the index measures total productivity change and not only shift. Since there are four candidates for the index, taking the geometric mean is a way of using all the information available (we will return to this in the next section). The index keeps its property of being an output index over an input index. This index is termed the Malmquist total factor productivity (MTFP) index in Bjurek (1996). It is interesting that Moorsteen

[2] Diewert also refers to Hicks (1961) who, according to Diewert, in a footnote perhaps suggested this type of index. Diewert named it the Hicks-Moorsteen approach. However, Moorsteen as well as Diewert only considered efficient operations, so the index was a technology shift index. Allowing for inefficiency as in Bjurek (1994) and (1996) the index will express total productivity change.

referred to the Paasche and the Laspeyre indexes when deriving the four productivity indexes without suggesting the geometric mean. The structure of the MTFP index as a geometric mean of four productivity indexes, all defined as ratios of an output quantity index and an input quantity index, is consistent with the structure of the Fisher Ideal index. The relationship between the MTFP index and the latter has not been exhaustively investigated. However, the corresponding relationship between the geometric mean of two Malmquist productivity indexes and the Fisher Ideal index has been discussed in Färe and Grosskopf (1992), Diewert (1992) and Balk (1993).

In this paper we compare the geometric mean formulations of the Malmquist productivity indexes and the Malmquist total factor productivity index given both constant and variable returns to scale technologies. In the variable returns to scale case the numerical value of the former index depends on whether it is input or output oriented. The MTFP index utilizes the input and the output orientation simultaneously. The types of indexes calculated and compared are shown in Table 5.1.

Scale property	Malmquist Productivity Index		Malmquist Total Factor Productivity Index
	Output-oriented	Input-oriented	
Constant returns	M(CRS)		MTFP(CRS)
Variable returns	M^o(VRS)	M^i(VRS)	MTFP(VRS)

Table 5.1 Types of productivity indexes computed

The paper is outlined as follows. In Section 5.2 the different indexes are introduced and a comparison of the indexes is carried out. In Section 5.3 we present the data and in Section 5.4 the empirical results are discussed. In Section 5.5 we summarize our findings.

5.2 THE MALMQUIST INDEXES

The Malmquist Productivity Index

As stated in the summary of CCD: "This paper develops index number procedures for making comparisons under very general circumstances. Malmquist input, output, and productivity comparisons are defined for structures of production with arbitrary returns to scale, substitution possibilities and biases in productivity change." (p. 1393). Moreover, they write: "There are two natural approaches to the measurement of productivity differences. One approach treats productivity differences as differences in maximum output conditional on a given level of inputs. This approach leads to *output based productivity indexes*. The alternative approach treats productivity differences as differences in minimum input requirement conditional on a given level of outputs. This view leads to *input based productivity indexes*. Output and input based productivity indexes differ from each other by a factor that reflects the returns to scale of the production structure." (pp. 1401-1402). The last statement was based on a discussion in Caves et al. (1981). The complexity of variable returns to scale in productivity indexes based on discrete data is discussed in Førsund (1996a, 1996b).

Let the technology for each period be represented by a *technology set*, S, defined as:

$$S = \{(x, y) : y \text{ can be produced by } x\} \qquad (5.2.1)$$

where y is the vector of M outputs and x the vector of N inputs. We will assume "standard" properties of this set, as e.g., found in Färe et al. (1994b), for example, S is closed, exhibits variable returns to scale, or constant returns to scale (meaning that a proportional scaling of a feasible point is also feasible), and free disposability of outputs and inputs. As is well known from the literature the output distance function is identical to the (extended[3]) Farrell out-

[3]The expression "extended" is used here because the reference technology may be a dif-

put increasing efficiency measure, $E^o(y,x)$[4]. The output distance function for a feasible point (x,y) is defined as:

$$D^o(y,x) = \min_{\delta}\left\{\delta : (x, \frac{y}{\delta}) \in S\right\} \equiv E^o(y,x), \delta > 0. \qquad (5.2.2)$$

Input distance functions may also be defined in a similar way, corresponding to the *inverse* of the (extended) Farrell input-saving efficiency measure, $E^i(y,x)$:

$$D^i(y,x) = \max_{\theta}\left\{\theta : (\frac{x}{\theta}, y) \in S\right\} \equiv \frac{1}{E^i(y,x)}, \theta > 0. \qquad (5.2.3)$$

Introducing observations from two periods, t and $t+1$, and corresponding technology sets S^t and S^{t+1}, the CCD definition of the Malmquist productivity index expressed in efficiency measures is:

$$M_s^k(y^t, x^t, y^{t+1}, x^{t+1}) = \frac{E_s^k(y^{t+1}, x^{t+1})}{E_s^k(y^t, x^t)}, s = t, t+1, k = o, i \quad (5.2.4)$$

One advantage of expressing the Malmquist index by the Farrell efficiency measures is that the output- and input-based index expressions are symmetric. A number greater than one signifies productivity improvement, and less than one productivity decline. In the following we will use the Farrell measure notation[5].

When measuring the productivity change between two periods the Malmquist index definition (5.2.4) gives us two measures corresponding to using the two technologies as reference technology. In

ferent one than the period of the observations, implying that the "cross period" efficiency scores may be larger than one.

[4]It is stated in, e.g., Färe et al. (1985), (1994b) that the Farrell efficiency measure is the inverse of the output distance function. However, Farrell (1957), p. 259, defined the output efficiency measure as positive and less or equal to one. Thus, the Farrell output efficiency measure is equal to the output distance function.

[5]In the definition of the quantity index Malmquist (1953, p. 230) introduces the nowdays well known idea of deflating or of a proportional scaling of all quantities such that "*For the quantity point $Q_1 = \{q_1\}$ a constant α is determined such that the quantity point $\alpha Q_1 = \{\alpha q_1\}$ is situated on [indifference] level S_0.*" This idea of Malmquist can obviously be formulated in terms of Farrell efficiency measures or distance functions as suggested by CCD.

Figure 5.1 we illustrate the Malmquist productivity indexes measured with respect to the technology at time $t+1$. Assume that point A represents a production unit at period t and point D the same unit at period $t+1$. The output based Malmquist productivity index is the ratio of the vertical line segments $x^{t+1}D/x^{t+1}H$ to the ratio $y^t/y^{F,t+1}$. The input based Malmquist productivity index is the ratio of the horizontal line segments $y^{t+1}G/y^{t+1}D$ to the ratio $x^{F,t+1}/x^t$. Since inefficiency is allowed the indexes do not only measure frontier shifts, i.e., technical change. It was shown in Färe et al. (1994a) that the indexes can be decomposed multiplicatively into a term measuring the relative technology shift, and one term measuring the catching-up of inefficient observations with their contemporaneous technology.

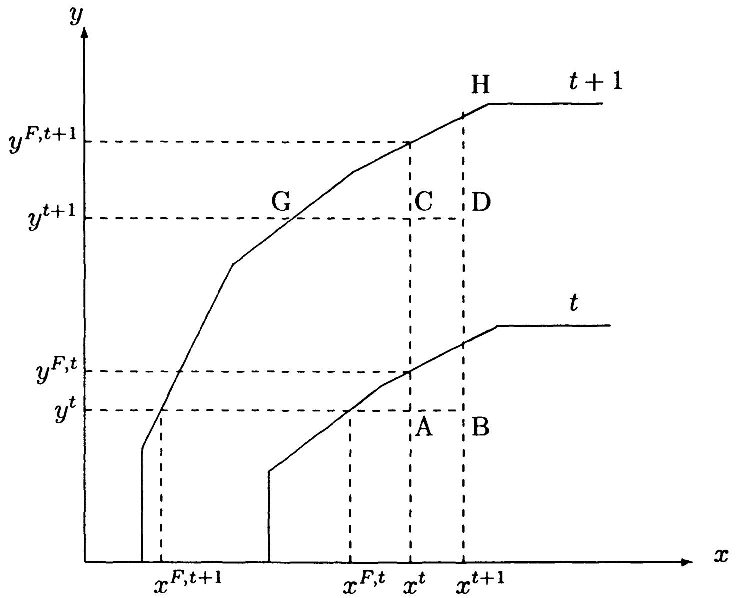

Figure 5.1 The Malmquist total factor productivity index.

In Figure 5.1 the technology shift in the output-oriented case of our example is measured by the relative distance between the frontiers at input level x^t. Using technology at period t as base, the shift term will be measured as the relative distance at level x^{t+1}. Without any strong reason to choose one distance for the other, a practical procedure is to take the geometric mean of the two distances. Indeed, Färe et al. (1994a) defined their Malmquist index as a geometric mean, and this practice is widespread now in the literature. We will therefore choose to report such geometric means:

$$M^k(y^t, x^t, y^{t+1}, x^{t+1}) = [M_t^k \cdot M_{t+1}^k]^{\frac{1}{2}} \left[\Pi_{s=t}^{t+1} \frac{E_s^k(y^{t+1}, x^{t+1})}{E_x^k(y^t, x^t)}\right]^{\frac{1}{2}},$$

$$k = o, i \tag{5.2.5}$$

The Malmquist Total Factor Productivity Index, MTFP

The definition of the Malmquist total factor productivity index, MTFP, for the production unit between periods $s = t$ and $t+1$, given the technologies at period $s = t$ and $t+1$ respectively, is defined as a ratio of an output quantity index and an input quantity index. These quantity index definitions are in correspondence with the approach in Malmquist (1953). The output quantity index is defined as:

$$MO_s(y^t, y^{t+1}, x^s) = \frac{E_s^o(y^{t+1}, x^s)}{E_s^o(y^t, x^s)}, s = t, t+1 \tag{5.2.6}$$

The definition of the input quantity index is:

$$MI_s(y^s, x^t, x^{t+1}) = \frac{E_s^i(y^s, x^t)}{E_s^i(y^s, x^{t+1})}, s = t, t+1 \tag{5.2.7}$$

The Malmquist output quantity index and the Malmquist input quantity index are defined in accordance with the definition in CCD. For instance, if MO is greater than unity, more outputs are produced at period $t+1$ than at period t, for the given technology at period s and quantity of inputs at period s. If MI is less than unity, fewer inputs have been used in production at period $t+1$

than at period t, for the given technology at period s and quantity of outputs produced at period s. The Malmquist total factor productivity index is defined as:

$$MTFP_s = \frac{MO_s(y^t, y^{t+1}, x^s)}{MI_s(y^s, x^t, x^{t+1})} = \frac{E_s^o(y^{t+1}, x^s)/E_s^o(y^t, x^s)}{E_s^i(y^s, x^t)/E_s^i(y^s, x^{t+1})}, s = t, t+1 \tag{5.2.8}$$

This index approach follows the common tradition of productivity indexes, i.e., it is a ratio between an output index and an input index. If the productivity index is greater than unity there is a productivity increase. If the index is less than unity, productivity has decreased and, if it is equal to unity, productivity is unchanged. Note that instead of defining the productivity index as an output-based or an input-based index this definition of the productivity index measures the change in output quantities in the output direction and the change in input quantities in the input direction, a property of particular importance given variable returns to scale technologies.

The productivity index (5.2.8) is the ratio of the output quantity index (5.2.6) to the input quantity index (5.2.7) measured with respect to the technology at period t or $t+1$. As mentioned in the introduction, two more productivity indexes can be calculated, each consisting of output and input quantity indexes calculated with respect to technologies at different periods; see Moorsteen (1961) and Diewert (1992). However, if we choose the geometric mean of the productivity indexes we obtain the same result if we calculate the geometric mean of all four possible productivity indexes or, as in (5.2.9) and in Bjurek (1994 and 1996), the geometric mean of the two indexes defined in (5.2.8). As in the case of the Malmquist productivity index, inefficiency is allowed in the Malmquist total factor productivity index and consequently it does not only measure technical change. The MTFP index reflects both technology shifts and changes in efficiency. The geometric mean maintains the fundamental structure of an output index to an input index. For a production unit observed at periods t and $t+1$, given the technologies at periods t and $t+1$ respectively it is defined as:

$$MTFP = \left[\Pi_{s=t}^{t+1} \frac{MO_s(y^t, y^{t+1}, x^s)}{MI_s(y^s, x^t, x^{t+1})}\right]^{\frac{1}{2}} \tag{5.2.9}$$
$$= \left[\Pi_{s=t}^{t+1} \frac{E_s^o(y^{t+1}, x^s)/E_s^o(y^t, x^s)}{E_s^i(y^s, x^t)/E_s^i(y^s, x^{t+1})}\right]^{\frac{1}{2}}$$

The construction of the MTFP index with respect to the technology at time $t+1$ can be illustrated in Figure 5.1, letting point A represent a production unit at period t and point D the same unit at period $t+1$. The output quantity index (5.2.6) in the Malmquist total factor productivity index is the ratio of the vertical line segments $x^{t+1}D/x^{t+1}H$ to $x^{t+1}B/x^{t+1}H$. The input quantity index (5.2.7) is the ratio of the horizontal line segments $y^{t+1}G/y^{t+1}C$ to $y^{t+1}G/y^{t+1}D$. Finally the MTFP$_{t+1}$ index (5.2.8) is the ratio of the output quantity index to the input quantity index.

A comparison among the Malmquist indexes

In this section we will briefly discuss the Malmquist productivity index and the Malmquist total factor productivity index. It is well-known that given strong disposability, piecewise linear variable returns to scale technologies, all efficiency measures (or distance functions) for a production unit included in the Malmquist productivity index may not be defined. This problem related to variable returns to scale may occur when efficiency measures are calculated for production units not belonging to the production set. This can be illustrated in Figure 5.1. Assume that we have the frontier technology at time t, and two production units observed at time $t+1$ represented by the input output combinations $(x^{F,t+1}, y^t)$ and (x^{t+1}, y^{t+1}). When the input quantity of a production unit at time $t+1$ is less than the minimum input quantity at time t, as for the unit $(x^{F,t+1}, y^t)$, the output efficiency measure is not defined. For the input efficiency measure the corresponding problem occurs for a unit at time $t+1$ with a larger output quantity than the maximum output quantity at time t, as for the unit (x^{t+1}, y^{t+1}). Though the problem has been known, the Malmquist productivity index has been applied in many studies given this technology assumption. A suggestion for overcoming this problem has been to introduce the two technologies, CRS and VRS, at the same time. The Malmquist index is calculated with respect to the CRS technology and the decomposition of the index is calculated with respect to the VRS technology, see Färe et al. (1994b). In contrast to the Malmquist productivity index, the problem of unbounded efficiency measures because of variable returns to scale technologies is not present in the Malmquist total factor productivity index; see Bjurek (1994 and 1996).

A second and more fundamental consequence of applying the Malmquist total factor productivity index is that this remedies the shortcoming of the Malmquist productivity index which measures total factor productivity change incorrectly due to variable returns to scale technologies. This problem of the Malmquist productivity index was most recently addressed in Grifell-Tatjé and Lovell (1995) and discussed when defining the Malmquist total factor productivity index in Bjurek (1996). The problem can be illustrated in Figure 5.1. Assume a case with no technical change represented by the technology $t + 1$. Moreover, assume that the input output combination $(x^{F,t+1}, y^t)$ represents the production unit at period t and the combination $(x^t, y^{F,t+1})$ the unit at period $t + 1$. The Malmquist productivity index measured with respect to the variable returns to scale technology shows that there is no productivity change since there is no frontier shift. However, it is obvious that productivity has decreased since the ratio of output to input for the unit in period t, $y^t/x^{F,t+1}$ is higher than the corresponding ratio for the unit in period $t+1$, $y^{F,t+1}/x^t$. Since the Malmquist total factor productivity index is a ratio of an output quantity index to an input quantity index, i.e., $MTFP = (y^{F,t+1}/y^t)/(x^t/x^{F,t+1})$, it consequently correctly reflects the productivity decrease[6].

[6]In the simple example with a single input and a single output the reader can easily verify that MTFP(VRS) = M(CRS). The CRS technology is represented by a straight line from the

The Malmquist productivity index and the Malmquist total factor productivity index generally do not coincide even under the assumption of constant returns to scale technologies (the former might be greater than, as well as, less than the latter). For instance they are not generally equal under the assumption of a nonparametric constant returns to scale technology with multiple outputs and multiple inputs. It can be shown that the indexes coincide if production units produce a single output with multiple inputs or if they produce multiple outputs with a single input given nonparametric constant returns to scale technologies. It can also be shown that they do coincide under constant returns to scale in the case of inverse homotheticity, see Färe et al. (1996), or in the case where all inputs and/or all outputs of a unit change proportionally. In the latter case the change in production is reduced to a situation of single input and/or single output.

5.3 DATA

The data used in this study comprise information on a large number of Swedish electricity retail distributors during a period of 21 years (1970-1990). Only distributors who supply more than 500 low voltage customers and employ more than 5 individuals are included. The total number of observations is 2275 constituting an unbalanced panel. The fact that 95% of the market for final consumption is supplied by two-thirds of the distributors, while the remaining third supplies only 5% of the customers, naturally affects the availability and quality of data on the smaller distributors. The data are constructed based on information obtained from the Association of Swedish Electric Utilities (SEF), Statistics Sweden (SCB) and different retail distributors.

---

origin through the most productive point on the technology $t + 1$ in Figure 5.1. In realistic cases with multiple inputs and/or multiple outputs and when variable returns to scale cannot be excluded a priori, the results are different if we measure productivity with respect to the VRS technology, compared to the results obtained if CRS is imposed in a similar way on the VRS technology.

As regards choice of output measure, we consider the total amount of low and high voltage electricity in MWhs received by the customers (Y_1, Y_2) and the number of low and high voltage customers (Y_3, Y_4) as the four outputs. On the input side we use kilometers of low and high voltage power lines (K_1, K_2) and total transformer capacity in kVa as the capital variables. Labour, L, is measured in full time equivalent employees.

Data for the period 1970-1987 has been utilized in Hjalmarsson and Veiderpass (1992) and for 1970-1990 in Kumbhakar and Hjalmarsson (forthcoming) for the same model specification, using the CCD definition of the Malmquist index in the case of CRS.

5.4 EMPIRICAL RESULTS

In this section we will illustrate the differences across the indexes introduced in Table 5.1. For the entire time period we have a complete panel of 29 units. However, for some periods the panel is much larger, e.g., 68 units 1970/71 and 134 units 1980/81.

In Figure 5.2 we have plotted the development of the geometric means of the Malmquist indexes for the panel units for the entire time period 1970-1990. To a surprisingly large extent the indexes almost coincide for large variations in productivity changes (amplitude ± 10%) and independent of definition and scale assumptions. This holds in particular for the two CRS indexes. Thus, the largest deviations are found for the VRS indexes, but among those, MTFP(VRS) is closer to the CRS measures than M^i(VRS) or M^o(VRS).

However, even if the CRS measures are close, behind the veil of the geometric means there is a lot of variation in individual units. This is revealed in Figure 5.3, where the ratios between MTFP(CRS) and M(CRS) are plotted for all consecutive years in the data set for the 29 panel units. Although most ratios are fairly close to one most periods there are some observations which deviate to some extent

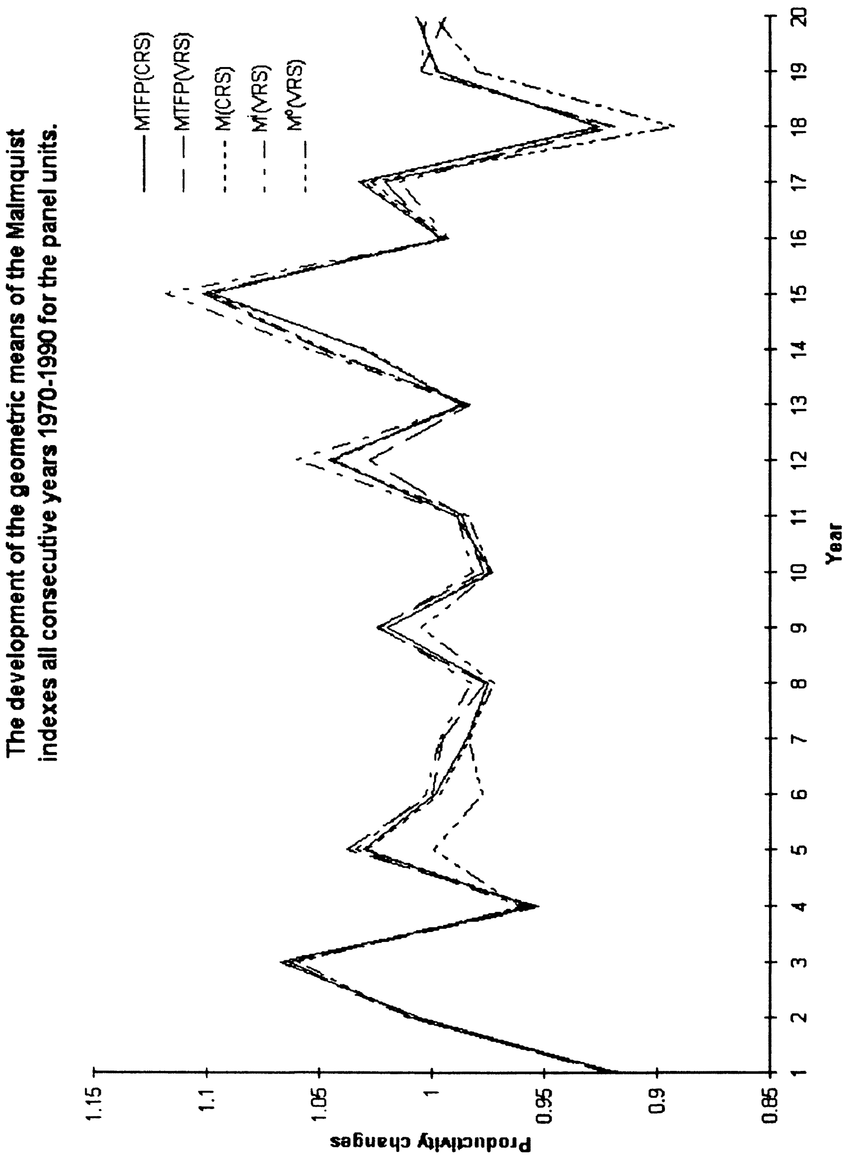
The development of the geometric means of the Malmquist indexes all consecutive years 1970-1990 for the panel units.
MTFP(CRS)
MTFP(VRS)
M(CRS)
M^i(VRS)
M^o(VRS)
Productivity changes
Year
1.15
1.1
1.05
1
0.95
0.9
0.85
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 5.2

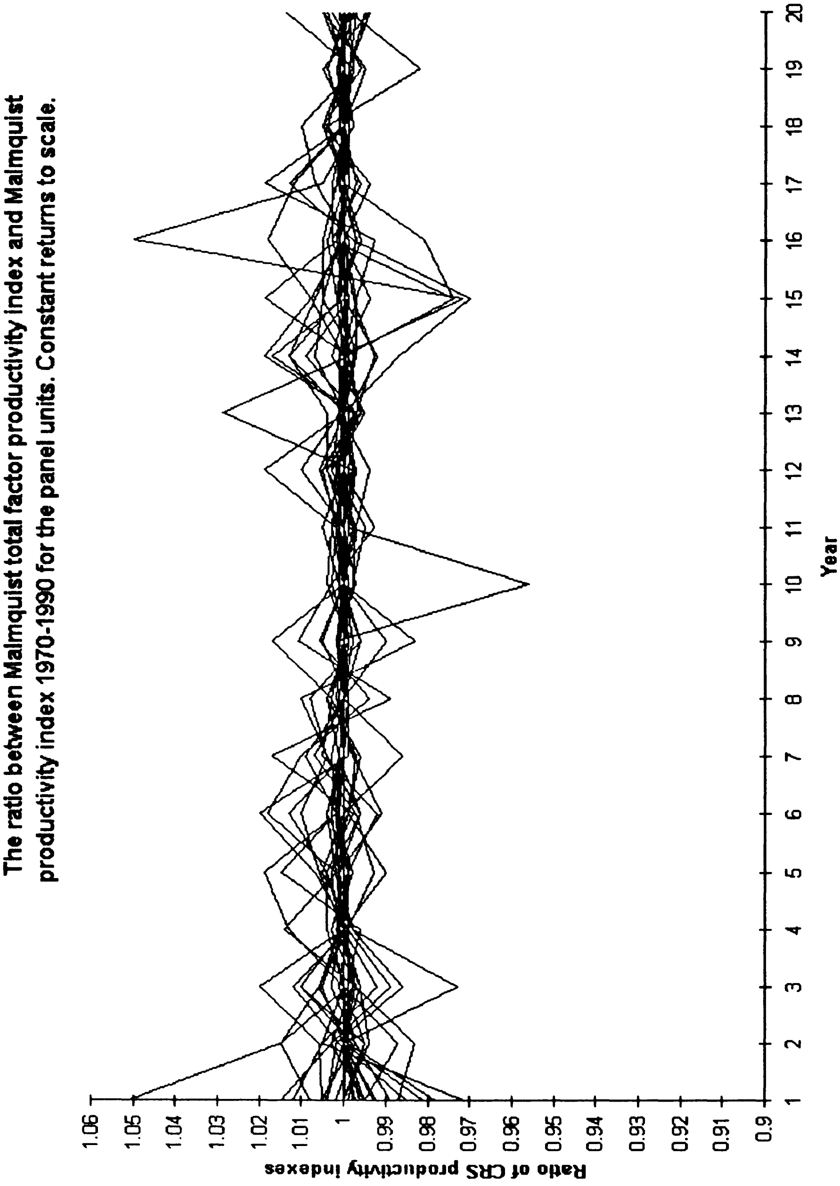
The ratio between Malmquist total factor productivity index and Malmquist productivity index 1970-1990 for the panel units. Constant returns to scale.
Ratio of CRS productivity indexes
1.06
1.05
1.04
1.03
1.02
1.01
1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
Year
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Figure 5.3

and a few which deviate considerably. All deviations from one are in the range of $\pm 5\%$. It is obvious from the figure that inverse homotheticity (i.e., simultaneous input and output homotheticity) does not hold at individual unit level.

In order to illustrate the index properties further, we have selected three units, one typical, one small and one large unit. Figure 5.4 shows the indexes for a typical unit in terms of size and development over time. Most periods the indexes are very close, but a few, 2-3 years, the differences are larger. While MTFP(VRS) and M^i(VRS) deviate upwards, all three VRS indexes deviate downwards from the CRS indexes.

Figure 5.5 illustrates the case with a typically small unit with a "noncomputability" problem, i.e., the LP solution degenerates so during the first 10 years M^o(VRS) does not exist because of small input values; see Figure 5.1 for an illustration of this case and the explanation on page 227. Here MTFP(VRS) deviates rather substantially from the other indexes during these 10 years, but thereafter it almost coincides with the other indexes, except M^o(VRS) which deviates somewhat from the other during the first years of its existence.

With reference to Figure 5.5 one might prematurely expect that VRS-measures are more volatile than CRS-measures. Figure 5.6 refutes this. It illustrates the case with a large unit with a nonexisting M^i(VRS) the first period because of large output values; see Figure 5.1 and the explanation on page 227. Here it turns out that the CRS-measures show the largest amplitude during the first periods, while M^i(VRS) exhibit this patter the later periods.

5.5 QUO VADIS, MALMQUIST?

The Malmquist productivity indexes proposed by CCD do not correctly represent change in total factor productivity in the case of non-constant returns to scale. This bias can be overcome by rein-

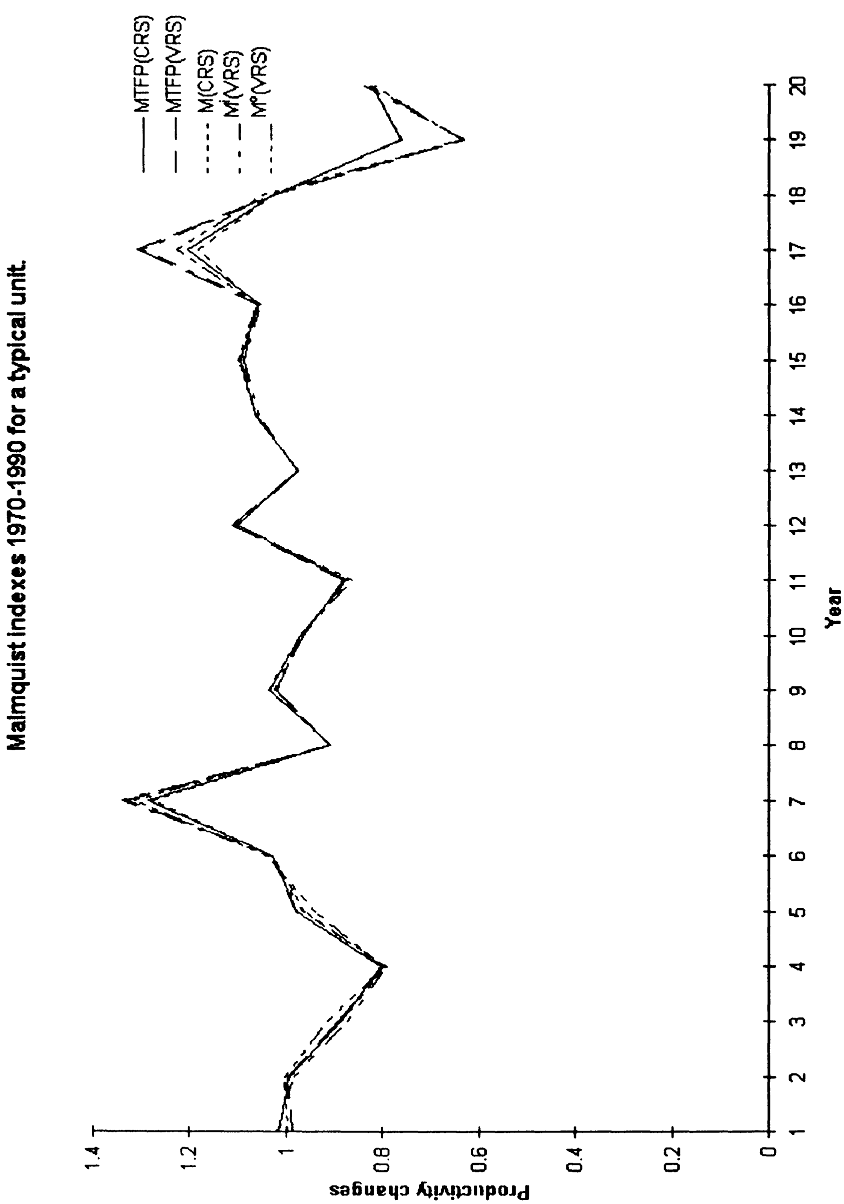
Malmquist indexes 1970-1990 for a typical unit.
MTFP(CRS)
MTFP(VRS)
M(CRS)
M^i(VRS)
M^o(VRS)
Productivity changes
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Year

Figure 5.4

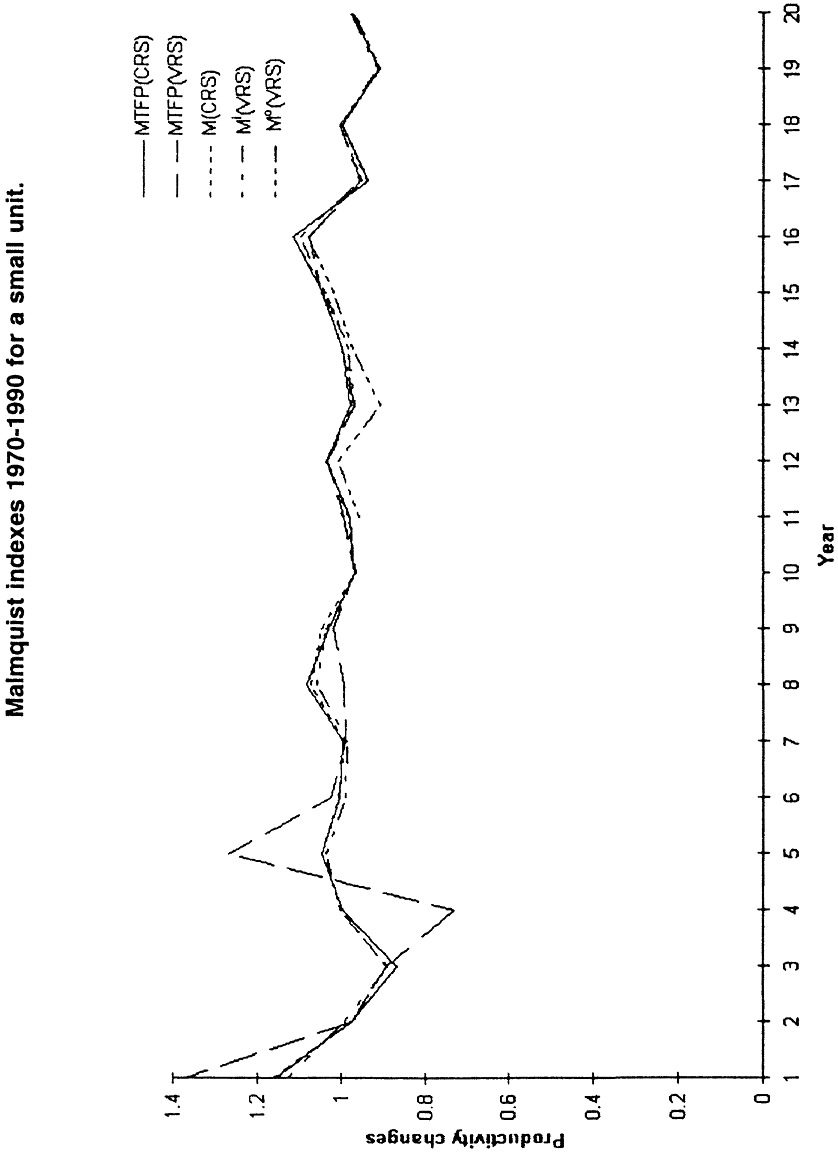
Malmquist indexes 1970-1990 for a small unit.
MTFP(CRS)
MTFP(VRS)
M(CRS)
M^I(VRS)
M^O(VRS)
Productivity changes
Year
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 5.5

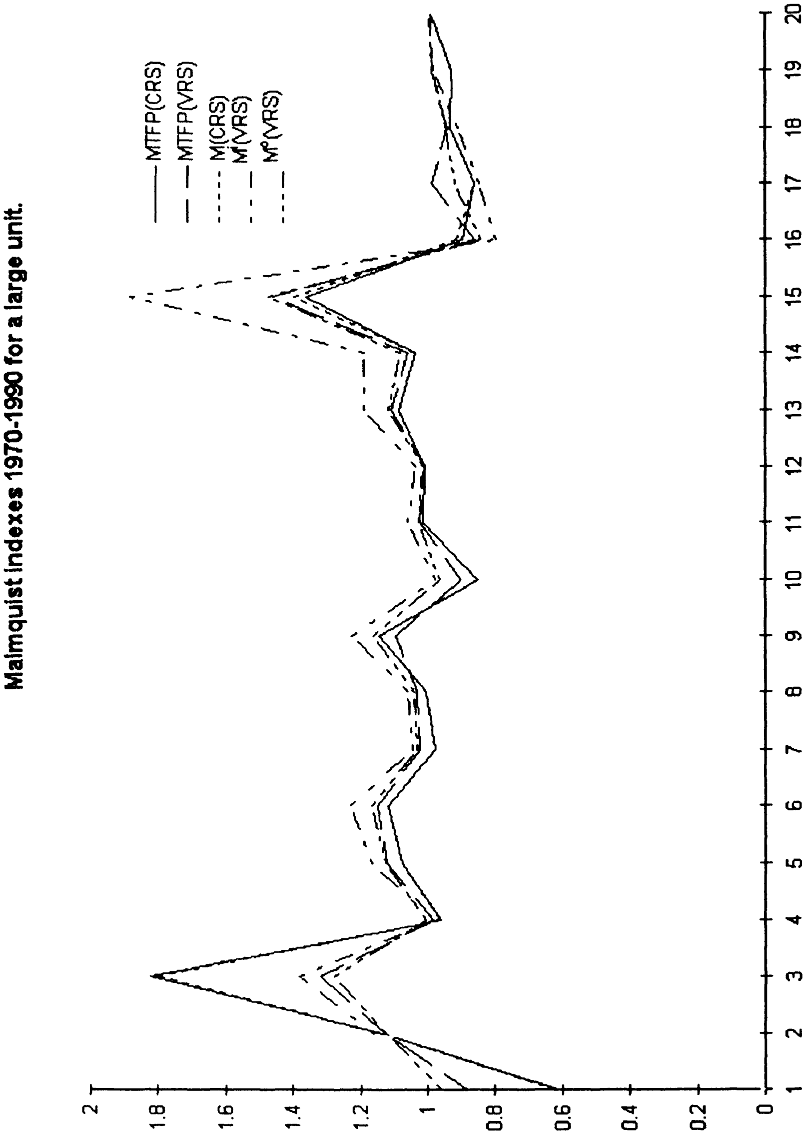
Malmquist indexes 1970-1990 for a large unit.
MTFP(CRS)
MTFP(VRS)
M(CRS)
M^{i}(VRS)
M^{o}(VRS)
Productivity changes
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Year

Figure 5.6

terpretation (see Førsund, 1996a), but there remains a problem of computation, as revealed in the discussion of Figure 5.1, and reflected in our results. The Malmquist total factor productivity index proposed by Bjurek (1996) is able to represent total factor productivity change for VRS technologies directly. Another attractive feature of the MTFP index is that its components always exist (under the standard conditions of no new goods or disappearing ones) in the case of variable returns to scale.

The empirical results highlight the following points:

a) If only mean values are of interest, the choice of index is not very important, and, thus, neither the choice between CRS and VRS.

b) For individual units the choice of index is of more importance in some years than others. However, a priori it is not possible to identify which units or indexes will deviate.

c) The productivity measures defined for VRS and CRS do not differ in volatility.

d) When specifying VRS, the Malmquist productivity indexes cannot always be computed for all units.

At this stage it is felt that it is too early to choose the ultimate Malmquist index. More empirical and theoretical effort is needed. One interesting issue for future research is the impact of non-homotheticity on the results, and the prevalence of non-homotheticity.

References

Balk, B.M. (1993) "Malmquist productivity indexes and Fisher Ideal indexes: comment," *Economic Journal* 103, 680-682.

Bjurek, H. (1994) "Essays on efficiency and productivity change with applications to public service production," *Ekonomiska Studier* 52, Department of economics, Göteborg University.

Bjurek, H. (1996) "The Malmquist total factor productivity index," *Scandinavian Journal of Economics*, 98(2), 303-313.

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.

Caves, D.W., L.R. Christensen, and J.A. Swanson (1981) "Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955-74," *American Economic Review*, 71, 994-1002.

Denny, M., M. Fuss and L. Waverman (1981) "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," in Cowing, T.G., and R.E. Stevenson (eds.), *Productivity Measurement in Regulated Industries*, New York: Academic Press, 179-218.

Diewert, W.E. (1981) "The Theory of Total Factor Productivity Measurement in Regulated Industries," in Cowing, T.G., and R.E. Stevenson (eds.), *Productivity Measurement in Regulated Industries*, New York: Academic Press, 17-44.

Diewert, W.E. (1992) "Fisher ideal output, input, and productivity indexes revisited," *Journal of Productivity Analysis*, 3, 211-248.

Farrell, M.J. (1957) "The measurement of produtive efficiency," *Journal of the Royal Statistical Society*, Series A (General), 120(III), 253-281(290).

Färe, R., and S. Grosskopf (1992) "Malmquist productivity indexes and Fisher Ideal indexes," *Economic Journal* 102, 158-160.

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

Nijhoff Publishing.

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

Färe, R., S. Grosskopf and P. Roos (1996) "On two Definitions of Productivity," *Economics Letters*, 55, 269-274.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1994a) "Productivity developments in Swedish hospitals: a Malmquist output index approach," in Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology, and Applications*, Boston/Dordrecht/London: Kluwer Academic Publishers, 253-272.

Førsund, F.R. (1996a) "The Malmquist productivity index, TFP and scale," memorandum no. 233, Department of Economics, University of Göteborg.

Førsund, F.R. (1996b) "Economic interpretations of the Malmquist productivity index," paper presented at Methodological Issues in Official Statistics Conference, Statistics Sweden, Stockholm.

Førsund, F.R., and L. Hjalmarsson (1987) Analyses of Industrial Structure. A Putty-Clay Approach. Stockholm: Almqvist & Wiksell International.

Grifell-Tatjé, E., and C.A.K. Lovell (1995) "A note on the Malmquist productivity index," *Economics Letters* 47, 169-175.

Haldi, J., and D. Whitcomb (1967) "Economies of Scale in Industrial Plants." *Journal of Political Economy*, 75, 373-385.

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, 18-31.

Hjalmarsson, L. and A. Veiderpass (1992) "Productivity in Swedish electricity retail distribution," *Scandinavian Journal of Economics*, 94, Supplement, 193-205.

Kumbhakar, S.C. and L. Hjalmarsson (forthcoming) "Relative performance of public and private ownership under yardstick regulation: Swedish electricity retail distribution, 1970-1990," *European Economic Review.*

Malmquist, S. (1953) "Index numbers and indifference surfaces," *Trabajos de Estadistica*, 4, 209-242.

Moorsteen, R.H. (1961) "On measuring productive potential and relative efficiency," *Quarterly Journal of Economics*, 75(3), 451-467.

Nishimizu, M. and J.M. Page (1982) "Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia 1965-78," *Economic Journal* 92, 920-936.

Pratten, C.F. (1971) "Economies of Scale in Manufacturing Industries." Department of applied economics, Occasional Papers, No. 28.

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

6

INPUT AND OUTPUT INDICATORS

by

Robert G. Chambers*

6.1 INTRODUCTION

This paper develops new input and output measures. The approach used to construct these measures relies on earlier work by Chambers (1996) that employed a version of Luenberger's (1992, 1995) shortage function to develop input, output, and productivity measures. The measures developed by Chambers (1996), being based on a translation representation of the technology, are to be contrasted directly with more conventional input and output measures which rely upon radial representations of the technology: namely, input and output distance functions (Caves, Christensen, and Diewert 1982a, 1982b).

In what follows, I first briefly discuss the translation measure upon which I base my new measures of inputs and outputs. Following Chambers, Chung, and Färe (1996), I refer to it as the *directional technology distance function* to emphasize that it represents a com-

*I would like to thank Bert Balk and Rolf Färe for many helpful comments.

plete generalization of Shephard's (1970) input and output distance functions. After briefly developing the properties of the directional technology distance function, I specify two parametric representations of it which are flexible in the sense of Diewert (1976). The first has the attractive property of automatically satisfying the translation properties of directional technology distance functions. I refer to this form as the *logarithmic-transcendental*. The second form is the quadratic, which was studied extensively in Chambers (1996). Next, I briefly discuss previous work on bilateral input, output, and productivity measurement and provide a synopsis of the main results in Chambers (1996) on Bennet-Bowley input and output measures. Then, I define new bilateral measures of input and output aggregates that are particularly appropriate for the logarithmic-transcendental technology and show how they can be calculated directly from observed data on input and output quantities and their prices. My last substantive section derives necessary and sufficient restrictions on the technology that insure that the bilateral input and output indicators defined here and by Chambers (1996) satisfy a form of additive transitivity that Blackorby and Donaldson (1980) refer to as additive circularity. These necessary and sufficient conditions are so restrictive that I then develop new multilateral input and output measures which satisfy additive circularity, but which can be constructed directly from a series of bilateral input and output indicators.

6.2 NOTATION, ASSUMPTIONS, AND DEFINITIONS

Let $\mathbf{x} \in \Re_+^N$ denote a vector of inputs and $\mathbf{y} \in \Re_+^M$ an output vector. Superscripts on input and output vectors are typically used to differentiate vectors either across time or across firms. (Exceptions are $\mathbf{0}^k$ and $\mathbf{1}^k$ which denote the k vectors of zeroes and ones, respectively.) For example, $\mathbf{x}^h$ will be interpreted variously as firm h's input use or as input use in period h. The technology is defined in terms of a set $T \subset \Re_+^N \times \Re_+^M$:

$$T = \{(\mathbf{x} \in \Re_{+}^{N}, \mathbf{y} \in \Re_{+}^{M}) : \mathbf{x} \text{ can produce } \mathbf{y}\}.$$

T satisfies the following properties:
T.1: T is closed;
T.2: Inputs and outputs are freely disposable, i.e., if $(\mathbf{x}', -\mathbf{y}') \geq (\mathbf{x}, -\mathbf{y})$ then $(\mathbf{x}, \mathbf{y}) \in T \Rightarrow (\mathbf{x}', \mathbf{y}') \in T$;
T.3: T is convex.

Slightly modifying Luenberger's (1992, 1995) shortage function and following Chambers, Chung, and Färe (1996), the *directional technology distance function* is defined by:

$$\begin{aligned}\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y) &= \sup\{\beta \in \Re : (\mathbf{x} - \beta\mathbf{g}_x, \mathbf{y} + \beta\mathbf{g}_y) \in T\}, \\ &\mathbf{g}_x \in \Re_{+}^{N}, \mathbf{g}_y \in \Re_{+}^{M}.\end{aligned}$$

where $(\mathbf{g}_x, \mathbf{g}_y) \neq (\mathbf{0}^n, \mathbf{0}^m)$. $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y)$ represents the maximal translation of the input and output vector in the direction of $(-\mathbf{g}_x, \mathbf{g}_y)$ that keeps the translated input and output vector inside T. When $(-\mathbf{g}_x, \mathbf{g}_y) = (-\mathbf{1}^n, \mathbf{1}^m)$, the directional technology distance function, therefore, is analogous to Blackorby and Donaldson's (1980) *translation function* for T. Figure 6.1 illustrates $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{1}^n, \mathbf{1}^m)$ as the ratio OB/OA for the point (x, y). All known distance and directional distance functions can be depicted as special cases of $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y)$. In particular, the *directional input distance function* defined by Chambers, Chung, and Färe (1995) is $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{0}^m)$, and the *directional output distance function* is $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{0}^n, \mathbf{g}_y)$.

An important property of the directional technology distance function is that it offers a complete function representation of the technology in that:

$$(\mathbf{x}, \mathbf{y}) \in T \Leftrightarrow \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y) \geq 0 \tag{6.2.1}$$

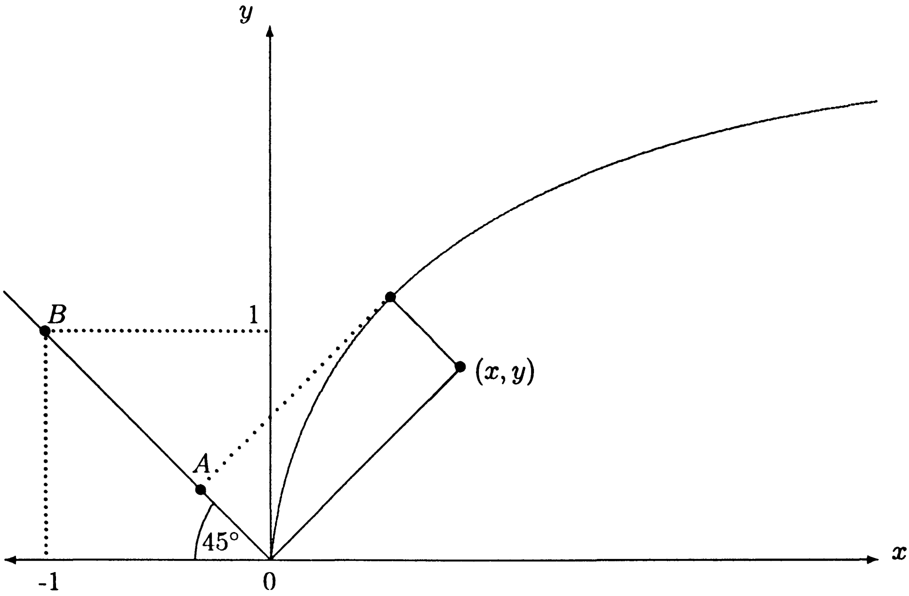

Figure 6.1 The Directional Technology Distance Function

(Luenberger, 1992). Points on the boundary of T are characterized by $\overrightarrow{D_T}(\mathbf{x},\mathbf{y};\mathbf{g}_x,\mathbf{g}_y) = 0$. Denote input prices by $\mathbf{w} \in \Re^n_{++}$ [2] and output prices by $\mathbf{p} \in \Re^m_{++}$. From (6.2.1), it follows immediately (Luenberger, 1992; Chambers, Chung, and Färe (1995, 1996)) that a profit maximizing firm solves:

$$\sup\left\{\mathbf{p}\cdot\left(\mathbf{y}+\overrightarrow{D_T}(\mathbf{x},\mathbf{y};\mathbf{g}_x,\mathbf{g}_y)\,\mathbf{g}_y\right)-\mathbf{w}\left(\mathbf{x}-\overrightarrow{D_T}(\mathbf{x},\mathbf{y}\mathbf{g}_x,\mathbf{g}_y)\,\mathbf{g}_x\right)\right\}. \tag{6.2.2}$$

Assuming that the directional technology distance function is differentiable in both inputs and outputs, the first-order conditions for an interior solution to (6.2.2) are:

$$\begin{aligned}\mathbf{p} &= -\nabla_y\overrightarrow{D_T}(\mathbf{x},\mathbf{y};\mathbf{g}_x,\mathbf{g}_y)\,(\mathbf{p}\cdot\mathbf{g}_y+\mathbf{w}\cdot\mathbf{g}_x) \\ \mathbf{w} &= \nabla_x\overrightarrow{D_T}(\mathbf{x},\mathbf{y};\mathbf{g}_x,\mathbf{g}_y)\,(\mathbf{p}\cdot\mathbf{g}_y+\mathbf{w}\cdot\mathbf{g}_x).\end{aligned} \tag{6.2.3}$$

[2] $\Re^k_{++}$ denotes the strictly positive k-orthant.

In equations (6.2.3) the notation ∇_z denotes the gradient of the function with respect to the vector $\mathbf{z}$.

The other properties of $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y)$ are summarized by Chambers, Chung, and Färe(1996):

D.1: $\overrightarrow{D_T}(\mathbf{x} - \alpha\mathbf{g}_x, \mathbf{y} + \alpha\mathbf{g}_y; \mathbf{g}_x, \mathbf{g}_y) = \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y) - \alpha, \alpha \in \Re$;

D.2: $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \lambda\mathbf{g}_x, \lambda\mathbf{g}_y) = \frac{1}{\lambda}\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y), \lambda > 0$;

D.3: $(\mathbf{x}', -\mathbf{y}') \geq (\mathbf{x}, -\mathbf{y}) \Rightarrow \overrightarrow{D_T}\left(\mathbf{x}', \mathbf{y}'; \mathbf{g}_x, \mathbf{g}_y\right) \geq \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y)$, i.e., nondecreasing in inputs and nonincreasing in output;

D.4: $\overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{g}_x, \mathbf{g}_y)$ is concave in $(\mathbf{x}, \mathbf{y})$ if T.3 is satisfied.

In what follows, I deal exclusively with a special case of the directional technology distance function (analogous to Blackorby and Donaldson's (1980) translation function), and I shall always deploy the more concise notation:

$$\begin{aligned} T_t(\mathbf{x}, \mathbf{y}) &\equiv \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{1}^n, \mathbf{1}^m), \\ T_i(\mathbf{x}, \mathbf{y}) &\equiv \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{1}^n, \mathbf{0}^m), \\ T_o(\mathbf{x}, \mathbf{y}) &\equiv \overrightarrow{D_T}(\mathbf{x}, \mathbf{y}; \mathbf{0}^n, \mathbf{1}^m), \end{aligned}$$

and refer to them, respectively, as the technology, input, and output translation functions.

The first parametric specification of the technology translation function that I consider is the *logarithmic-transcendental.* Firm h's technology translation function is logarithmic-transcendental if it can be written in the form:

$$\exp T_t^h(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}^h \exp(\frac{x_i}{2}) \exp(\frac{x_j}{2})$$

$$+\frac{1}{2}\sum_{k=1}^{m}\sum_{j=1}^{m} b_{jk}^{h}\exp(\frac{-y_j}{2})\exp(\frac{-y_k}{2})$$
$$+\sum_{j=1}^{n}\sum_{k=1}^{m} c_{kj}^{h}\exp(\frac{x_j}{2})\exp(\frac{-y_k}{2}),$$

where $a_{ij}^h = a_{ji}^h$ for all i and j, $b_{jk}^h = b_{kj}^h$ for j and k, and $c_{kj}^h = c_{jk}^h$ for all j and k. The logarithmic-transcendental, being a member of the generalized-quadratic class of functions (Blackorby, Primont, and Russell, 1978), is second-order flexible. Also notice that the logarithmic-transcendental form automatically satisfies property D.1 for the technology translation function.

The quadratic[3] k translation function (k= i,o) for firm h is:

$$T_k^h(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} a_i^h x_i + \sum_{k=1}^{m} b_k^h y_k + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}^h x_i x_j$$
$$+\frac{1}{2}\sum_{k=1}^{m}\sum_{l=1}^{m}\beta_{kl}^h y_k y_l + \sum_{i=1}^{n}\sum_{k=1}^{m}\gamma_{ik}^h x_i y_k,$$

with $\alpha_{ij}^h = \alpha_{ji}^h, \beta_{kl}^h = \beta_{lk}^h$. If this form is interpreted as an input-translation function, the following parametric restrictions insure that D.1 is satisfied:

$$\sum_{i=1}^{n} a_i^h = 1; \sum_{j=1}^{n}\alpha_{ij}^h = 0, i = 1, ..., n; \sum_{i=1}^{n}\gamma_{ik}^h = 0, k = 1, ..., m.$$

And finally if the quadratic is interpreted as an output translation function then the following insure that D.1 is satisfied:

$$\sum_{k=1}^{m} b_k^h = -1; \sum_{l=1}^{m}\beta_{kl}^h = 0, k = 1, ..., m; \sum_{k=1}^{m}\gamma_{ik}^h = 0, i = 1, .., n.$$

[3]Notice that the logarithmic-transcendental has fewer parameters than the quadratic. However, the logarithmic-transcendental automatically satisfies D.1, while the quadratic does not. To see how the quadratic must be further restricted, differentiate D.1, $T_t(\mathbf{x} - \alpha\mathbf{1}^n, \mathbf{y} + \alpha\mathbf{1}^m) = T_t(\mathbf{x},\mathbf{y}) - \alpha$ with respect to α to obtain $-\nabla_{\mathbf{x}} T_t(\mathbf{x} - \alpha\mathbf{1}^n, \mathbf{y} + \alpha\mathbf{1}^m)\cdot\mathbf{1}^n + \nabla_{\mathbf{y}} T_t(\mathbf{x} - \alpha\mathbf{1}^n, \mathbf{y} + \alpha\mathbf{1}^m)\cdot\mathbf{1}^m = -1$. Now differentiate this expression with respect to α and evaluate both expressions at $\alpha = 1$.

6.3 PREVIOUS WORK ON INPUT AND OUTPUT INDEXES

Input and output indexes are summary measures of two things: multiple input use or multiple output production and input and output differences either over time, place, or firms. For the purposes of our present discussion, suppose that one is considering how a firm's use of a single input, which we denote as x, changes over time. There are at least two ways to measure this change: The first and perhaps the most obvious is simply the difference between input use in period one, x^1, and input use in the base period, x^0, i.e., $x^1 - x^0$. This approach brings with it the advantage that changes in the origin from which these two numbers are measured has no effect on the measure of input change, but it also has the disadvantage that the input-change measure is not unit free. For example, if we go from measuring the input in terms of hours to measuring in terms of hundreds of hours, the input measure changes. A second approach, which remedies the unit problem, is to consider the ratio of input use in period one to input use in the base period, i.e., x^1/x^0. However, this ratio-approach has the disadvantage that changes in the origin from which these two numbers are measured does have an effect on the resulting measure of input change. So, for example, if we are originally measuring input committal in hours worked and then move to measuring input committal in terms of hours over 5 hours worked, the ratio measure must typically change, and in some instances the measure will not even be well defined. To see this clearly, consider the case where x^0 was originally 5 hours worked, and the translation of the origin described above takes place. The new index is not well-defined.

In fact, one of the most common empirical problems that occurs in applying ratio-based indexes is what to do with zero observations, as ratio-based indexes are typically not well defined at the origin. Difference based indexes typically will be well-defined at the origin precisely because they are invariant to changes in the origin.

Index numbers have almost exclusively been calculated using the ratio approach. All the traditionally familiar indexes (Laspeyres, Paasche, and Fisher's ideal) are ratio-based measures. Moreover, these more traditional indexes were all calculated using a "test" or axiomatic approach to index-number construction. That is, tests reflecting reasonable properties that an index should possess were specified, and indexes meeting these tests were then derived. A more recent development has been the derivation of indexes using what Samuelson and Swamy (1974) have referred to as the economic approach to index numbers. In the economic approach to index numbers, indexes are constructed from primal representations of preferences and technology under the presumption that individual agents are economic optimizers. Following the original work of Konüs (1939) and Malmquist (1953), virtually all of these indexes have been calculated using a ratio approach. Here the basic idea is relatively simple: Take a radial measure of the technology, a distance function, and then express input and output indexes in terms of ratios of these measures of the technology. Because radial functional representations of technology are positively linearly homogeneous, these economic measures are invariant to the units in which quantities are calculated.

Perhaps the most influential works in this area have been the papers by Caves, Christensen, and Diewert (1982a, 1982b) which show that the Törnqvist approximation to the Divisia index is an exact index that can be derived by taking geometric averages of Malmquist indexes for transcendental logarithmic input and output distance functions. Because the transcendental-logarithmic function is second-order flexible, these results imply that the Törnqvist index is 'superlative' in Diewert's (1976) sense.

The only studies, to my knowledge, which have pursued economic indexes of input and outputs using the difference approach are Diewert (1992), Diewert (1993) and Chambers (1996). Diewert (1993) briefly discusses Bennet's (1920) index as a possible way of accounting for changes in inputs or outputs but does not develop any such indexes in detail. Although raised in a different context, the early works of Bennet (1920) on measuring the cost of living and Bowley (1928) on welfare evaluation are direct predecessors of this difference approach. In fact, Chambers (1996a) shows that the Bennet-Bowley index is an exact measure of Allais' (1943) disposable surplus when the consumer's benefit function is quadratic.

I now briefly survey the results by Chambers (1996) on input and output measurement. As a starting point, it is convenient to define two measures of profitability. The *cost-based measure* for input prices $\mathbf{w}$ and input levels $\mathbf{x^1}$ and $\mathbf{x^0}$ is:

$$C\left(\mathbf{w};\mathbf{x}^1,\mathbf{x}^0\right)=\mathbf{w}\cdot\left(\mathbf{x}^1-\mathbf{x}^0\right).$$

The *revenue-based measure* for output prices $\mathbf{p}$ and output bundles $\mathbf{y}^1$ and $\mathbf{y}^0$ is:

$$R\left(\mathbf{p};\mathbf{y}^1,\mathbf{y}^0\right)=\mathbf{p}\cdot(y^1-\mathbf{y}^0).$$

Depending upon where input prices are evaluated, $C\left(\mathbf{w};\mathbf{x}^1,\mathbf{x}^0\right)$ and $R\left(\mathbf{p};\mathbf{y}^1,\mathbf{y}^0\right)$ are the analogues in difference form of the Laspeyres or Paasche input and output indexes.

The *Bennet-Bowley cost-based measure* is the average of the Laspeyres and Paasche cost-based measures:

$$BC(\mathbf{w}^1,\mathbf{w}^0;\mathbf{x}^1,\mathbf{x}^0)=\frac{1}{2}\left(C\left(\mathbf{w}^1;\mathbf{x}^1,\mathbf{x}^0\right)+C\left(\mathbf{w}^0;\mathbf{x}^1,\mathbf{x}^0\right)\right).$$

The *Bennet-Bowley revenue-based measure* is:

$$BR(\mathbf{p}^1,\mathbf{p}^0;\mathbf{y}^1,\mathbf{y}^0)=\frac{1}{2}\left(R\left(\mathbf{p}^1;\mathbf{y}^1,\mathbf{y}^0\right)+R\left(\mathbf{p}^0;\mathbf{y}^1,\mathbf{y}^0\right)\right).$$

The Bennet-Bowley measures, of course, are the difference analogues of the appropriate Fisher ideal indexes. Notice, however, that they also have the attractive intuitive property that they can be interpreted as cost differences, revenue differences, and profit differences evaluated at average prices (e.g., the Bennet-Bowley cost-based measure measures the difference in costliness of the two input bundles $\mathbf{x}^1$ and $\mathbf{x}^0$ at average input prices $\frac{1}{2}\left(\mathbf{w}^0+\mathbf{w}^1\right)$).

Chambers (1996) defines the *1-technology Luenberger input indicator* for $(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1)$ by:

$$X^1\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1\right) = T_i^1\left(\mathbf{x}^0, \mathbf{y}^1\right) - T_i^1\left(\mathbf{x}^1, \mathbf{y}^1\right),$$

and the 0-technology Luenberger input indicator for $(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1)$ by:

$$X^0\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0\right) = T_i^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - T_i^0\left(\mathbf{x}^1, \mathbf{y}^0\right).$$

Figure 6.2 illustrates $X^1(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1)$ in the case where firms operate inefficiently as the difference between the amounts that one can translate $\mathbf{x}^0$ and $\mathbf{x}^1$ in the direction of the bisector and still keep both input bundles in the input set for technology 1. In the case illustrated, $X^1(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1) > 0$, suggesting that $\mathbf{x}^0$ is *larger* than $\mathbf{x}^1$ when the difference in the input bundles is measured relative to the ability to produce $\mathbf{y}^1$ using technology 1. On the other hand, $X^0(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0)$ compares the two input bundles' ability to produce $\mathbf{y}^0$ relative to technology 0. It would be desirable to have an indicator that is invariant to the technology chosen to make the comparison. A natural compromise is to take the average of these two indicators.

The *Luenberger input indicator*, denoted $X(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1)$, is defined:

$$X\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right) = \frac{1}{2}\left(X^1\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1\right) + X^0\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0\right)\right).$$

An obvious consequence of these definitions and D.1 (the translation property) is

(6.3.4) **Theorem**: *(Chambers, 1996)* $X^k(\mathbf{x}^0 - \alpha\mathbf{1}^n, \mathbf{x}^1 - \alpha\mathbf{1}^n, \mathbf{y}^k) = X^k\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^k\right)$, $k = 0, 1$.

(6.3.5) **Corollary**: *(Chambers, 1996)* $X(\mathbf{x}^0 - \alpha\mathbf{1}^n, \mathbf{x}^1 - \alpha\mathbf{1}^n, \mathbf{y}^1, \mathbf{y}^0) = X(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1)$.

Put in words, the theorem and the corollary say that all the Luenberger input indicators are *translation invariant in inputs.* This

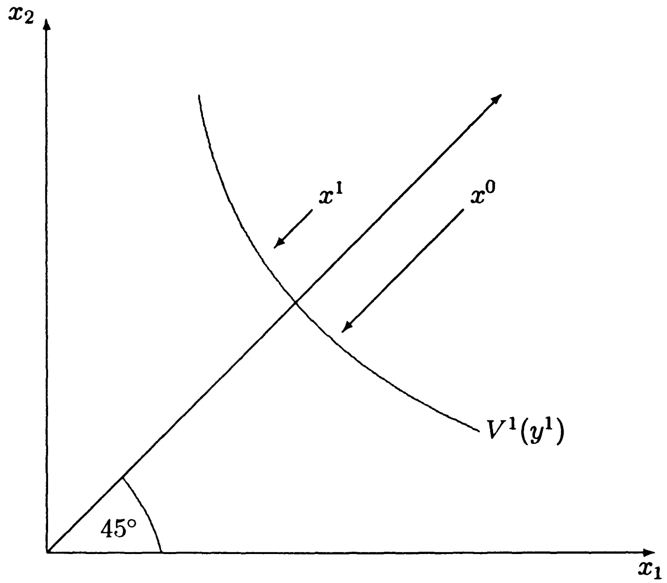

Figure 6.2 $X^1(x^0, x^1, y^1)$

should be contrasted directly with Malmquist input indexes' homogeneity of degree zero in inputs. In the case of Malmquist indexes, zero degree homogeneity emerges from the linear homogeneity of input distance functions in inputs. Here, translation invariance follows from D.1. Effectively, it makes the difference between input aggregates independent of the choice of the origin. Chambers (1996) main result on Luenberger input indicators is:

(6.3.6) **Theorem**: *(Chambers, 1996)* If the firm minimizes cost, the input-translation function is quadratic with $\alpha_{ij}^0 = \alpha_{ij}^1$ for all i and j, then:

$$X\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right) = -BC(\tilde{\mathbf{w}}^1, \tilde{\mathbf{w}}^0; \mathbf{x}^1, \mathbf{x}^0).$$

where $\tilde{\mathbf{w}}^k = \frac{\mathbf{w}^k}{\mathbf{w}^k \cdot \mathbf{1}^n}$.

This theorem is important because it implies that the Bennet-Bowley cost-based measure is an exact indicator for a second-order flexible technology. (Notice, as Caves, Christensen, and Diewert (1982) point out, that the restriction $\alpha_{ij}^0 = \alpha_{ij}^1$ restricts the flexibility of the technology.) Hence, the Bennet-Bowley cost-based measure might be thought of as a superlative indicator of input differences.

Parallel to the definition of the input indicators, Chambers (1996) defines the *1-technology Luenberger output indicator* for $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{y}^0)$ by:

$$Y^1\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1\right) = T_o^1\left(\mathbf{x}^1, \mathbf{y}^0\right) - T_o^1\left(\mathbf{x}^1, \mathbf{y}^1\right),$$

and the 0-technology Luenberger output indicator for $(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0)$ by:

$$Y^0\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0\right) = T_o^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - T_o^0\left(\mathbf{x}^0, \mathbf{y}^1\right).$$

$Y^k\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^k\right)$ thus measures the difference between the amounts $\mathbf{y}^0$ and $\mathbf{y}^1$ can be projected in the direction of the bisector and still keep both of them in the $\mathbf{x}^k$ output set for technology k. The *Luenberger output indicator* is the average of $Y^1(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1)$ and $Y^0(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0)$:

$$Y\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right) = \frac{1}{2}\left(Y^1\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1\right) + Y^0\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0\right)\right).$$

An obvious consequence of these definitions and D.1 is

(6.3.7) **Theorem**: *(Chambers, 1996)* $Y^k\left(\mathbf{y}^0 + \alpha\mathbf{1}^m, \mathbf{y}^1 + \alpha\mathbf{1}^m, \mathbf{x}^k\right) = Y^k\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^k\right), k = 0, 1.$

(6.3.8) **Corollary**: *(Chambers, 1996)* $Y\left(\mathbf{y}^0 + \alpha\mathbf{1}^m, \mathbf{y}^1 + \alpha\mathbf{1}^m, \mathbf{x}^0, \mathbf{x}^1\right) = Y\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right).$

Chambers (1996) also shows that the Bennet-Bowley revenue based measure is an exact measure (and thus superlative) of the Luenberger output indicator under appropriate assumptions on the technology:

(6.3.9) **Theorem**: *(Chambers, 1996)* If the firm maximizes revenue, the output translation function is quadratic with $\beta_{ij}^0 = \beta_{ij}^1$ for all i and j, then

$$Y\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right) = BR(\tilde{\mathbf{p}}^1, \tilde{\mathbf{p}}^0; \mathbf{y}^1, \mathbf{y}^0),$$

where $\tilde{\mathbf{p}}^k = \frac{\mathbf{p}^k}{\mathbf{p}^k \cdot \mathbf{1}^m}$.

6.4 EXPONENTIAL INPUT AND OUTPUT INDICATORS

Where Chambers (1996) defines indicators in terms of input and output translation function, here I want to define input and output indicators in terms of the logarithmic-transcendental technology translation function. To that end, I define the *1-technology exponential input indicator* for $(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1)$ by:

$$EX^1\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1\right) = \exp T_t^1\left(\mathbf{x}^0, \mathbf{y}^1\right) - \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right),$$

and the 0-technology exponential input indicator for $(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1)$ by:

$$EX^0\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0\right) = \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - \exp T_t^0\left(\mathbf{x}^1, \mathbf{y}^0\right).$$

The primary difference between the exponential input indicators introduced here and the indicators studied in Chambers (1996) is that these indicators are specified in terms of differences of exponentials of technology translation functions, while those in Chambers (1996) are specified in terms of differences of input translation functions. However, as with the Luenberger indicators defined by Chambers (1996), both compare the ability of the two input bundles to produce different output bundles relative to a different technology.

The *Luenberger exponential input indicator* is the simple average of the 1-technology and 0-technology exponential input indicators, i.e.,

$$EX\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right) = \frac{1}{2}\left(EX^1\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^1\right) + EX^0\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0\right)\right).$$

An obvious consequence of these definitions and D.1 (the translation property) is

(6.4.1) **Theorem**: $EX^k\left(\mathbf{x}^0 - \alpha\mathbf{1}^n, \mathbf{x}^1 - \alpha\mathbf{1}^n, \mathbf{y}^k + \alpha\mathbf{1}^m\right) = \exp(-\alpha)\, EX^k\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^k\right)$,
$k = 0, 1, \alpha \in \Re$

(6.4.2) **Corollary**: $EX\left(\mathbf{x}^0 - \alpha\mathbf{1}^n, \mathbf{x}^1 - \alpha\mathbf{1}^n, \mathbf{y}^0 + \alpha\mathbf{1}^m, \mathbf{y}^1 + \alpha\mathbf{1}^m\right) = \exp(-\alpha)\, EX\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right)$,
$\alpha \in \Re$.

I define the 1-*technology exponential output indicator* for $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{y}^0)$ by:

$$EY^1\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1\right) = \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^0\right) - \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right),$$

and the 0-technology exponential output indicator for $(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0)$ by:

$$EY^0\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0\right) = \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^1\right).$$

$EY^k\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^k\right)$ thus measures the difference between the amounts $\mathbf{y}^0$ and $\mathbf{y}^1$ can be projected in the direction of the bisector and still keep both of them in the $\mathbf{x}^k$ output set for technology k . The *Luenberger exponential output indicator* is the average of $EY^1\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1\right)$ and $EY^0\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0\right)$:

$$EY\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right) = \frac{1}{2}\left(EY^1\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^1\right) + EY^0\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}0\right)\right)$$

An obvious consequence of these definitions and D.1 is

(6.4.3) **Theorem**: $EY^k\left(\mathbf{y}^0 + \alpha\mathbf{1}^m, \mathbf{y}^1 + \alpha\mathbf{1}^m, \mathbf{x}^k - \alpha\mathbf{1}^n\right) = \exp(-\alpha) EY^k\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^k\right)$,
$k = 0, 1$.

(6.4.4) **Corollary**: $EY\left(\mathbf{y}^0 + \alpha\mathbf{1}^m, \mathbf{y}^1 + \alpha\mathbf{1}^m, \mathbf{x}^0 - \alpha\mathbf{1}^n, \mathbf{x}^1 - \alpha\mathbf{1}^n\right) = \exp(-\alpha) EY\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right)$.

6.5 EXACT MEASURES OF THE LUENBERGER EXPONENTIAL INPUT AND OUTPUT INDICATORS

This section derives exact measures of the exponential input and output indicators introduced in the previous section that can be calculated without econometric estimation. My first result is:

(6.5.1) **Theorem**: If the technology translation function is logarithmic-transcendental with $a_{ij}^0 = a_{ij}^1$, for all i and j, and firms maximize profit, then:

$$EX\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right) = \sum_{k=1}^{n} \left(\exp\left(\frac{-x_k^0}{2}\right) \bar{w}_k^0 + \exp\left(\frac{-x_k^1}{2}\right) \bar{w}_k^1\right) \left(\exp\left(\frac{x_k^0}{2}\right) - \exp\left(\frac{x_k^1}{2}\right)\right),$$

where $\bar{w}_k^h = \frac{w_k^h}{\mathbf{w}^h \cdot \mathbf{1}^n + \mathbf{p}^h \cdot \mathbf{1}^m}, h = 0, 1.$

Proof By Diewert's (1976) quadratic identity:

$$\exp T_t^1\left(\mathbf{x}^0, \mathbf{y}^1\right) - \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right) = \frac{1}{2}\left[\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^0, \mathbf{y}^1\right) + \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right)\right] \cdot \left[\exp\left(\frac{\mathbf{x}^0}{2}\right) - \exp\left(\frac{\mathbf{x}^1}{2}\right)\right].$$

Similarly,

$$\exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - \exp T_t^0\left(\mathbf{x}^1, \mathbf{y}^0\right) = \frac{1}{2}\left[\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) + \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^1, \mathbf{y}^0\right)\right] \cdot \left[\exp\left(\frac{\mathbf{x}^0}{2}\right) - \exp\left(\frac{\mathbf{x}^1}{2}\right)\right]$$

Adding these two equalities gives:

$$\left[\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0,\mathbf{y}^0\right) + \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1,\mathbf{y}^1\right)\right] \cdot \left[\exp\left(\frac{\mathbf{x}^0}{2}\right) - \exp\left(\frac{\mathbf{x}^1}{2}\right)\right] +$$
$$\frac{1}{2}\left[\begin{array}{c} \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^0,\mathbf{y}^1\right) + \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^1,\mathbf{y}^0\right) \\ -\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0,\mathbf{y}^0\right) - \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1,\mathbf{y}^1\right) \end{array}\right]$$
$$\cdot\left[\exp\left(\frac{\mathbf{x}^0}{2}\right) - \exp\left(\frac{\mathbf{x}^1}{2}\right)\right]$$

Under the assumption that $a_{ij}^0 = a_{ij}^1$ for all i and j the second term in this expression is zero. Hence,

$$EX\left(\mathbf{x}^0,\mathbf{x}^1,\mathbf{y}^0,\mathbf{y}^1\right) = \frac{1}{2}\left[\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0,\mathbf{y}^0\right) + \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1,\mathbf{y}^1\right)\right] \cdot\left[\exp\left(\frac{\mathbf{x}^0}{2}\right) - \exp\left(\frac{\mathbf{x}^1}{2}\right)\right] \tag{6.5.2}$$

Differentiation establishes that:

$$\exp T_t(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}} T_t(\mathbf{x},\mathbf{y}) = \frac{1}{2}\exp\left(\frac{\mathbf{x}}{2}\right) \circ \nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t\left(\mathbf{x},\mathbf{y}\right)$$

where $\mathbf{z} \circ \mathbf{v}$ denotes the vector consisting of the component-wise products of the vectors $\mathbf{z}$ and $\mathbf{x}$. When the firm maximizes profit, this last expression upon using (6.2.3) reduces to

$$\exp\left(-\frac{\mathbf{x}}{2}\right) \circ \bar{\mathbf{w}} = \frac{1}{2}\nabla_{\exp\left(\frac{\mathbf{x}}{2}\right)} \exp T_t\left(\mathbf{x},\mathbf{y}\right).$$

This last expression when substituted into (6.5.2) establishes the theorem.

A parallel argument establishes:

(6.5.3) **Theorem**: If the technology translation function is logarithmic-transcendental with $b_{ij}^0 = b_{ij}^1$ for all i and j and firms maximize profit, then:

$$EY\left(\mathbf{y}^0, \mathbf{y}^1, \mathbf{x}^0, \mathbf{x}^1\right) = \sum_{k=1}^{m} \left[\exp\left(\frac{y_k^0}{2}\right)\bar{p}_k^0 + \exp\left(\frac{y_k^1}{2}\right)\bar{p}_k^1\right] \left[\exp\left(\frac{-y_k^0}{2}\right) - \exp\left(\frac{-y_k^1}{2}\right)\right].$$

where $\bar{p}_k^h = \frac{p_k^h}{\mathbf{p}^h \cdot \mathbf{1}^m + \mathbf{w}^h \cdot \mathbf{1}^n}, h = 0, 1.$

Proof By Diewert's (1976) identity:

$$\exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^0\right) - \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right) = \frac{1}{2}\left[\nabla_{\exp\left(\frac{-bmy}{2}\right)} \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^0\right) + \nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right)\right] \cdot \left[\exp\left(-\frac{\mathbf{y}^0}{2}\right) - \exp\left(\frac{-\mathbf{y}^1}{2}\right)\right].$$

Similarly,

$$\exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) - \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^1\right) = \frac{1}{2}\left[\nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) + \nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^1\right)\right] \cdot \left[\exp\left(\frac{-\mathbf{y}^0}{2}\right) - \exp\left(\frac{-\mathbf{y}^1}{2}\right)\right].$$

Adding these two expressions together and rearranging under the assumption that $b_{ij}^0 = b_{ij}^1$ gives:

$$\left[\nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t^0\left(\mathbf{x}^0, \mathbf{y}^0\right) + \nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t^1\left(\mathbf{x}^1, \mathbf{y}^1\right)\right] \cdot \left[\exp\left(\frac{-\mathbf{y}^0}{2}\right) - \exp\left(\frac{-\mathbf{y}^1}{2}\right)\right]$$

Differentiation establishes:

$$\exp T_t(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{y}} T_t(\mathbf{x},\mathbf{y}) = -\frac{1}{2}\exp\left(-\frac{\mathbf{y}}{2}\right) \circ \nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t(\mathbf{x},\mathbf{y}).$$

Together with (6.2.3), this last result establishes for a profit maximizing firm that:

$$\frac{1}{2}\nabla_{\exp\left(\frac{-\mathbf{y}}{2}\right)} \exp T_t(\mathbf{x},\mathbf{y}) = \bar{\mathbf{p}} \circ \exp\left(\frac{\mathbf{y}}{2}\right).$$

The result follows immediately.

6.6 TRANSITIVITY OF INPUT AND OUTPUT INDICATORS

So far, all the indicators studied only make bilateral comparisons either across firms or time. More generally, however, one will want to make multilateral comparisons. For example, one potential application of the input and output indicators is for the construction of a time series of an aggregate input from multiple time series of single inputs. A property that is usually deemed desirable in the construction of multilateral indexes is Frisch circularity. Here the analogue of Frisch circularity corresponds to what Blackorby and Donaldson (1980) have referred to as *additive circularity*. A multilateral indicator $G(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k)$ satisfies additive circularity if:

$$G(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k) = G(\mathbf{x}^h, \mathbf{x}^j, \mathbf{y}^h, \mathbf{y}^j) + G(\mathbf{x}^j, \mathbf{x}^k, \mathbf{y}^j, \mathbf{y}^k)$$

for all h, j, and k. From this definition, it is apparent why additive circularity might be a desirable property for an indicator to possess. Consider using a multilateral indicator $G(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k)$ to construct a series of observations on aggregate inputs across firms. Let the observation on the base firm's (denoted now by a superscript 1) aggregate input be normalized to G^1. Notice, for example, that there are at least two ways to construct the aggregate input for firm 3: First, one can construct it directly by using the formula: $G^3 = G^1 + G(\mathbf{x}^3, \mathbf{x}^1, \mathbf{y}^3, \mathbf{y}^1)$. Or, one could construct it more indirectly by first constructing G^2, and then computing $G^2 + G(\mathbf{x}^3, \mathbf{x}^2, \mathbf{y}^3, \mathbf{y}^2)$. Unless the indicator satisfies additive circularity, there is no reason for the result of both computations to be the same.

The reason, of course, that this happens is that the choice of a base observation here is essentially arbitrary. Generally speaking, when looking at comparisons across firms (or across countries for that matter), there is no natural way of ranking firms. However, in some applications, there is a natural ordering, and when there is, additive circularity becomes a less compelling property to possess. A clear example of this is in the construction of a time series of aggregate inputs. In that case, the ordering is clear, and one can usefully construct meaningful aggregates by making successive bilateral comparisons from the base period.[4]

When additive circularity is important it comes at a severe cost in terms of restricting the technologies to which it implies. My first result of this section is that:

(6.6.1) **Theorem** The Luenberger input indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if:

$$T_i^h(\mathbf{x}, \mathbf{y}) = a^h(\mathbf{y}) + b(\mathbf{x}),$$

where $b(\mathbf{x} + \alpha \mathbf{1}^n) = b(\mathbf{x}) + \alpha$.

Proof By Lemma 1 in Blackorby and Donaldson (1980), the Luenberger input indicator satisfies additive circularity if and only if it can be expressed as:

$$X\left(\mathbf{x}^0, \mathbf{x}^1, \mathbf{y}^0, \mathbf{y}^1\right) = v\left(\mathbf{x}^0, \mathbf{y}^0\right) - v\left(\mathbf{x}^1, \mathbf{y}^1\right)$$

[4] In the literature on time-series aggregates, this is referred to as chain linking.

where $v(\mathbf{x}^0 + \alpha \mathbf{1}^n, \mathbf{y}^0) = v(\mathbf{x}^0, \mathbf{y}^0) + k\alpha$. Substituting into the definition of the Luenberger input indicator and setting $\mathbf{x}^1 = \mathbf{0}^n$ and $\mathbf{y}^1 = \mathbf{0}^m$ gives:

$$\begin{aligned} T_i^0\left(\mathbf{x}^0, \mathbf{y}^0\right) &= T_i^0\left(\mathbf{0}^n, \mathbf{y}^0\right) + 2v\left(\mathbf{x}^0, \mathbf{y}^0\right) - 2v\left(\mathbf{0}^n, \mathbf{0}^m\right) \\ &\quad + T_i^1\left(\mathbf{x}^0, \mathbf{0}^m\right) - T_i^1\left(\mathbf{0}^n, \mathbf{0}^m\right). \end{aligned}$$

Performing a similar operation for $X\left(\mathbf{x}^0, \mathbf{x}^h, \mathbf{y}^0, \mathbf{y}^h\right)$ establishes that:

$$T_i^1\left(\mathbf{x}^0, \mathbf{0}^m\right) - T_i^1\left(\mathbf{0}^n, \mathbf{0}^m\right) = T_i^h\left(\mathbf{x}^0, \mathbf{0}^m\right) - T_i^h\left(\mathbf{0}^n, \mathbf{0}^m\right),$$

so that we can rewrite the above in the obvious renormalization:

$$T_i^0\left(\mathbf{x}^0, \mathbf{y}^0\right) = t_i^0\left(\mathbf{y}^0\right) + m\left(\mathbf{x}^0, \mathbf{y}^0\right).$$

Because this same argument can be applied for arbitrary h and k it follows immediately that:

$$T_i^h(\mathbf{x}, \mathbf{y}) = t_i^h(\mathbf{y}) + m(\mathbf{x}, \mathbf{y}). \tag{6.6.2}$$

where $m(\mathbf{x}, \mathbf{y})$ must satisfy D.1 in $\mathbf{x}$. Substituting this result into the additive circularity condition and simplifying yields:

$$\begin{aligned} -m\left(\mathbf{x}^1, \mathbf{y}^0\right) + m\left(\mathbf{x}^0, \mathbf{y}^1\right) - m\left(\mathbf{x}^2, \mathbf{y}^1\right) + m\left(\mathbf{x}^1, \mathbf{y}^2\right) &= \\ -m\left(\mathbf{x}^2, \mathbf{y}^0\right) + m\left(\mathbf{x}^0, \mathbf{y}^2\right) & \end{aligned}$$

Now set $\mathbf{x}^1 = \mathbf{x}^2 = \mathbf{0}^n, \mathbf{y}^1 = \mathbf{y}^0 = \mathbf{0}^m$ to obtain:

$$m\left(\mathbf{x}^0, \mathbf{0}^m\right) - m\left(\mathbf{0}^n, \mathbf{0}^m\right) + m\left(\mathbf{0}^n, \mathbf{y}^2\right) = m\left(\mathbf{x}^0, \mathbf{y}^2\right)$$

so that:

$$m(\mathbf{x}, \mathbf{y}) = b(\mathbf{x}) + n(\mathbf{y})$$

in an obvious renormalization. This result along with (6.6.2) establishes necessity. Sufficiency is obvious.

(6.6.3) **Corollary**: If the Luenberger input indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$, then it is independent of output.

Hence, transitivity in the form of additive circularity places severe restrictions on the classes of technology which will permit one to construct meaningful Luenberger input indicators: All firms must possess a technology whose input translation function is additively separable in inputs and outputs, and where only the function dealing with outputs can be specific to firms. Some intuitive insight into the form that this technology assumes can be had by noticing that applying D.1 to the form in the theorem yields:

$$T_i^h(\mathbf{x},\mathbf{y}) = a^h(\mathbf{y}) + b(\mathbf{x}) = b(\mathbf{x} + a^h(\mathbf{y}) \cdot \mathbf{1}^n).$$

So transitive Luenberger input indicators are only available if the technology differences across firms can be summarized by a common input-translation function that is independent of the level of output, $b(\mathbf{x})$, and actual differences emerge as a result of a firm-specific translation of inputs along the unit vector, where the magnitude of the translation depends upon the output choice. In a sense, the differences across firms are restricted to changing the efficiency with which a given vector of inputs is utilized. Figure 6.3 illustrates: The base technology (h=1) is represented by the input set characterized by the isoquant labelled V^1. The technology for h=2 is found by translating every element of the base isoquant by $a^2(\mathbf{y}) - a^1(\mathbf{y})$ along the unit vector.

Turning to the quadratic input translation function, it follows immediately that:

(6.6.4) **Corollary**: If the input-translation function is quadratic, it satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if $a_i^h = a_i^k$ for all h and k and $i = 1, 2, ..., n, \alpha_{ij}^h = \alpha_{ij}^k$ for all h and k and $i, j = 1, 2, ..., n$, and $\gamma_{ik}^h = 0$ for all h, i, k.

An exactly parallel argument establishes:

(6.6.5) **Theorem**: The Luenberger output indicator satisfies

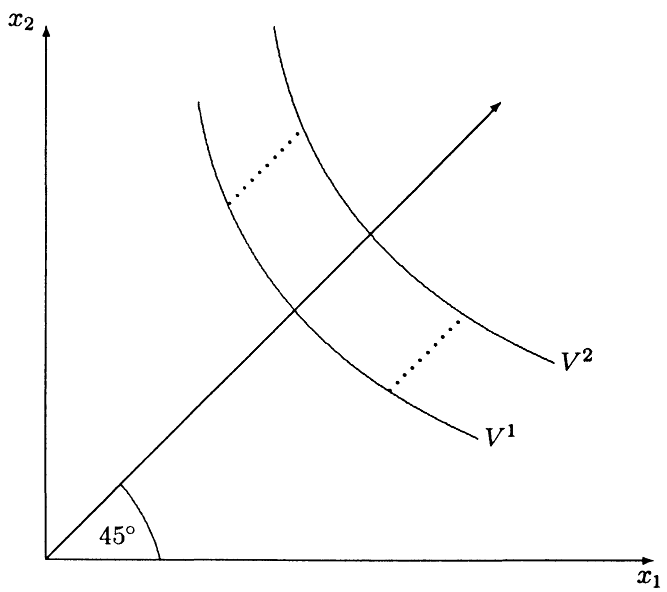

Figure 6.3 Additive Circularity

additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if:

$$T_o^h(\mathbf{x}, \mathbf{y}) = a(\mathbf{y}) + b^h(\mathbf{x}),$$

where $a(\mathbf{y} + \beta \mathbf{1}^m) = a(\mathbf{y}) - \beta$ for all h.

(6.6.6) **Corollary**: If the Luenberger output indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$, it is independent of inputs.

(6.6.7) **Corollary**: If the output-translation function is quadratic, it satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if $b_k^h = b_k^j$ for all h and j, $k = 1, 2, ..m, \beta_{kl}^h = \beta_{kl}^j$ for all h and j, $k, l = 1, 2, ..., m$, and $\gamma_{ik}^h = 0$ for all h, i, k.

So for Luenberger output indicators to be transitive it must be true that the technology can be described as though there exists a single reference output set common across firms with the only effect that inputs have on production to be in terms of translating outputs, by a firm-specific amount that depends on the input mix, along the unit vector.

Similar results apply for the Luenberger exponential indicators:

(6.6.8) **Theorem**: The Luenberger exponential input indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if:

$$\exp T_t^h(\mathbf{x}, \mathbf{y}) = a^h(\mathbf{y}) + b(\mathbf{x}),$$

where $a^h(\mathbf{y} + \beta \mathbf{1}^m) + b(\mathbf{x} - \beta \mathbf{1}^n) = \exp(-\beta) \left(a^h(\mathbf{y}) + b(\mathbf{x})\right)$ for all h.

(6.6.9) **Corollary**: If the Luenberger exponential input indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$, it is independent of output.

(6.6.10) **Corollary**: If the technology translation function is logarithmic-transcendental, it satisfies additive circularity of the Luenberger exponential input indicator if and only if: $a_{ij}^h = a_{ij}^t$ all h and t, $i, j = 1, 2, ..., n$, and $c_{kj}^h = 0$ for all h.

(6.6.11) **Theorem**: The Luenberger exponential output indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ if and only if:

$$\exp T_t^h(\mathbf{x}, \mathbf{y}) = a(\mathbf{y}) + b^h(\mathbf{x}),$$

where $a(\mathbf{y} + \beta \mathbf{1}^m) + b^h(\mathbf{x} - \beta \mathbf{1}^n) = \exp(-\beta)(a(\mathbf{y}) + b^h(\mathbf{x}))$ for all h.

(6.6.12) **Corollary**: If the Luenberger exponential output indicator satisfies additive circularity for all $\mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ it is independent of inputs.

(6.6.13) **Corollary**: If the technology translation function is logarithmic-transcendental, it satisfies additive circularity of the Luenberger exponential output indicator if and only if: $b_{ij}^h = \hat{b}_{ij}^t$ all h and t, $i, j = 1, 2, ..., m$, and $c_{kj}^h = 0$ for all h.

The primary implication of the preceding theorems and their corollaries is that it is highly unlikely that any technology will satisfy the restrictions required for additive circularity of the bilateral indicators to hold. Hence, using the bilateral approach that we have developed so far will not be sufficient to ensure the existence of multilateral input and output indicators that satisfy this form of additive transitivity. However, by suitably redefining the indicators previously developed, it is possible to generate both multilateral input and output indicators that satisfy additive circularity. Following Caves, Christensen, and Diewert (1982a, 1982b), we consider two separate approaches: In terms of constructing input indicators, the first approach is to take an arbitrary observation on inputs and outputs, call it $(\mathbf{x}^*, \mathbf{y}^*)$, and to construct all bilateral input indicators relative to this observation, i.e., create as the case warrants either $X\left(\mathbf{x}^h, \mathbf{x}^*, \mathbf{y}^h, \mathbf{y}^*\right)$ or $EX\left(\mathbf{x}^h, \mathbf{x}^*, \mathbf{y}^h, \mathbf{y}^*\right)$. Once these indicators relative to a common base are constructed, new indicators expressed relative to this common base can be defined as:

$$\vec{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right) = X\left(\mathbf{x}^h, \mathbf{x}^*, \mathbf{y}^h, \mathbf{y}^*\right) - X\left(\mathbf{x}^k, \mathbf{x}^*, \mathbf{y}^k, \mathbf{y}^*\right),$$

and

$$E\vec{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right) = EX\left(\mathbf{x}^h, \mathbf{x}^*, \mathbf{y}^h, \mathbf{y}^*\right) - EX\left(\mathbf{x}^k, \mathbf{x}^*, \mathbf{y}^k, \mathbf{y}^*\right).$$

It is easy to verify that these new indicators are, in fact, transitive and satisfy the additive circularity property.

Another alternative is to make all input comparisons relative to the average technology by taking the average of either $X\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right)$ or $EX\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right)$, as appropriate, over all possible k, i.e., define new indicators:

$$\bar{X}(\mathbf{x}^h, \mathbf{y}^h) = \frac{1}{N}\sum_{k=1}^{N} X\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right),$$

and

$$E\bar{X}(\mathbf{x}^h, \mathbf{y}^h) = \frac{1}{N}\sum_{i=1}^{N} EX\left(\mathbf{x}^h, \mathbf{x}^i, \mathbf{y}^h, \mathbf{y}^i\right).$$

$\bar{X}(\mathbf{x}^h, \mathbf{y}^h)$ and $E\bar{X}(\mathbf{x}^h, \mathbf{y}^h)$ can be thought of as indicators of input usage for firm h relative to the average of input usage by the firms considered. Once these average indicators are constructed new bilateral indicators can then be derived as the difference between these average indicators. That is, as

$$\hat{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right) = \bar{X}(\mathbf{x}^h, \mathbf{y}^h) - \bar{X}(\mathbf{x}^k, \mathbf{y}^k),$$

and

$$E\hat{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right) = E\bar{X}(\mathbf{x}^h, \mathbf{y}^h) - E\bar{X}(\mathbf{x}^k, \mathbf{y}^k).$$

Both $\hat{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right)$ and $E\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right)$ satisfy additive circularity, and are thus transitive.

We thus have,

(6.6.14) **Theorem**: If the firm minimizes cost, the input-translation function is quadratic with $\alpha_{ij}^h = \alpha_{ij}^k$ for all h and k, then:

$$\begin{aligned}\vec{X}\left(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k\right) &= BC(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^*; \mathbf{x}^k, \mathbf{x}^*) \\ &\quad - BC(\tilde{\mathbf{w}}^*, \tilde{\mathbf{w}}^h; \mathbf{x}^*, \mathbf{x}^h),\end{aligned}$$

and

$$\begin{aligned}\hat{X}(\mathbf{x}^h, \mathbf{x}^k, \mathbf{y}^h, \mathbf{y}^k) &= -\frac{1}{N}\sum_{i=1}^{N}(BC(\tilde{\mathbf{w}}^i, \tilde{\mathbf{w}}^h; \mathbf{x}^i, \mathbf{x}^h) \\ &\quad - BC(\tilde{\mathbf{w}}^i, \tilde{\mathbf{w}}^k; \mathbf{x}^i, \mathbf{x}^k))\end{aligned}$$

where $\tilde{\mathbf{w}}^k = \frac{\mathbf{w}^k}{\mathbf{w}^k \cdot \mathbf{1}^n}$.

Similarly,

(6.6.15) **Theorem**: If the technology translation function is logarithmic-transcendental with $a^h_{ij} = a^k_{ij}$ for all h and k, and firms maximize profit, then:

$$\begin{aligned} E\vec{X}\left(\mathbf{x}^h,\mathbf{x}^k,\mathbf{y}^h,\mathbf{y}^k\right) = & \sum_{k=1}^{n}\left(\exp\left(\frac{-x^h_k}{2}\right)\bar{w}^h_k + \exp\left(\frac{-x^*_k}{2}\right)\bar{w}^*_k\right) \\ & \left(\exp\left(\frac{x^h_k}{2}\right) - \exp\left(\frac{x^*_k}{2}\right)\right) \\ & + \sum_{k=1}^{n}\left(\exp\left(\frac{-x^*_k}{2}\right)\bar{w}^*_k + \exp\left(\frac{-x^k_k}{2}\right)\bar{w}^k_k\right) \\ & \left(\exp\left(\frac{x^*_k}{2}\right) - \exp\left(\frac{x^k_k}{2}\right)\right), \end{aligned}$$

and

$$\begin{aligned} E\hat{X}\left(\mathbf{x}^h,\mathbf{x}^k,\mathbf{y}^h,\mathbf{y}^k\right) = & \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{n}\left(\exp\left(\frac{-x^h_k}{2}\right)\bar{w}^h_k + \exp\left(\frac{-x^i_k}{2}\right)\bar{w}^i_k\right) \\ & \left(\exp\left(\frac{x^h_k}{2}\right) - \exp\left(\frac{x^i_k}{2}\right)\right) \\ & - \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{n}\left(\exp\left(\frac{-x^k_k}{2}\right)\bar{w}^k_k + \exp\left(\frac{-x^i_k}{2}\right)\bar{w}^i_k\right) \\ & \left(\exp\left(\frac{x^k_k}{2}\right) - \exp\left(\frac{x^i_k}{2}\right)\right) \end{aligned}$$

where $\bar{w}^h_k = \frac{w^h_k}{\mathbf{w}^h \cdot \mathbf{1}^n + \mathbf{p}^h \cdot \mathbf{1}^m}$.

Construction of output indicators satisfying additive circularity follows a similar procedure: Either define a common base against which all outputs are compared, or use the average output indicator as the common base. Therefore, we have:

$$\vec{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) = Y\left(\mathbf{y}^h,\mathbf{y}^*,\mathbf{x}^h,\mathbf{x}^*\right) - Y\left(\mathbf{y}^k,\mathbf{y}^*,\mathbf{x}^k,\mathbf{x}^*\right),$$

$$E\vec{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) = EY\left(\mathbf{y}^h,\mathbf{y}^*,\mathbf{x}^h,\mathbf{x}^*\right) - EY\left(\mathbf{y}^k,\mathbf{y}^*,\mathbf{x}^k,\mathbf{x}^*\right),$$

$$\begin{aligned}\hat{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) &= \frac{1}{N}\sum_{j=1}^{N} Y\left(\mathbf{y}^h,\mathbf{y}^j,\mathbf{x}^h,\mathbf{x}^j\right)\\ &\quad -\frac{1}{N}\sum_{j=1}^{N} Y\left(\mathbf{y}^k,\mathbf{y}^j,\mathbf{x}^k,\mathbf{x}^j\right),\end{aligned}$$

and

$$\begin{aligned}E\hat{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) &= \frac{1}{N}\sum_{j=1}^{N} EY\left(\mathbf{y}^h,\mathbf{y}^j,\mathbf{x}^h,\mathbf{x}^j\right)\\ &\quad -\frac{1}{N}\sum_{j=1}^{N} EY\left(\mathbf{y}^k,\mathbf{y}^j,\mathbf{x}^k,\mathbf{x}^j\right).\end{aligned}$$

Each of these new indicators, which we shall refer to as *Luenberger multilateral output indicators and Luenberger exponential multilateral output indicators*, respectively, satisfy additive circularity, and we obtain as before:

(6.6.16) **Theorem**: If firms maximize revenue, the output translation function is quadratic with $\beta_{ij}^m = \beta_{ij}^n$ for all m and n, then

$$\begin{aligned}\vec{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) &= BR(\tilde{\mathbf{p}}^*,\tilde{\mathbf{p}}^h;\mathbf{y}^h,\mathbf{y}^*)\\ &\quad -BR(\tilde{\mathbf{p}}^*,\tilde{\mathbf{p}}^k;\mathbf{y}^k,\mathbf{y}^*),\end{aligned}$$

and

$$\hat{Y}\left(\mathbf{y}^h,\mathbf{y}^k,\mathbf{x}^h,\mathbf{x}^k\right) = \frac{1}{N}\sum_{j=1}^{N} BR(\tilde{\mathbf{p}}^h,\tilde{\mathbf{p}}^j;\mathbf{y}^j,\mathbf{y}^h)$$

$$-\frac{1}{N}\sum_{j=1}^{N} BR(\tilde{\mathbf{p}}^j, \tilde{\mathbf{p}}^k; \mathbf{y}^j, \mathbf{y}^k),$$

where $\tilde{\mathbf{p}}^k = \frac{\mathbf{p}^k}{\mathbf{p}^k \cdot \mathbf{1}^m}$.

(6.6.17) **Theorem**: If the technology translation function is logarithmic-transcendental with $b_{ij}^m = b_{ij}^n$ for all m and n, and firms maximize profit, then:

$$\begin{aligned} E\vec{Y}\left(\mathbf{y}^h, \mathbf{y}^k, \mathbf{x}^h, \mathbf{x}^k\right) = & \sum_{k=1}^{m}\left[\exp\left(\frac{y_k^h}{2}\right)\bar{p}_k^h + \exp\left(\frac{y_k^*}{2}\right)\bar{p}_k^*\right] \\ & \left[\exp\left(\frac{-y_k^h}{2}\right) - \exp\left(\frac{-y_k^*}{2}\right)\right] \\ & -.\sum_{k=1}^{m}\left[\exp\left(\frac{y_k^k}{2}\right)\bar{p}_k^1 + \exp\left(\frac{y_k^*}{2}\right)\bar{p}_k^*\right] \\ & \left[\exp\left(\frac{-y_k^k}{2}\right) - \exp\left(\frac{-y_k^*}{2}\right)\right], \end{aligned}$$

and

$$\begin{aligned} E\hat{Y}\left(\mathbf{y}^h, \mathbf{y}^k, \mathbf{x}^h, \mathbf{x}^k\right) = & \sum_{j=1}^{N}\sum_{k=1}^{m}\left[\exp\left(\frac{y_k^h}{2}\right)\bar{p}_k^h + \exp\left(\frac{y_k^j}{2}\right)\bar{p}_k^i\right] \\ & \left[\exp\left(\frac{-y_k^h}{2}\right) - \exp\left(\frac{-y_k^j}{2}\right)\right] \\ & -.\sum_{j=1}^{N}\sum_{k=1}^{m}\left[\exp\left(\frac{y_k^k}{2}\right)\bar{p}_k^k + \exp\left(\frac{y_k^j}{2}\right)\bar{p}_k^j\right] \\ & \left[\exp\left(\frac{-y_k^k}{2}\right) - \exp\left(\frac{-y_k^j}{2}\right)\right] \end{aligned}$$

where $\bar{p}_k^h = \frac{p_k^h}{\mathbf{p}^h \cdot \mathbf{1}^m + \mathbf{w}^h \cdot \mathbf{1}^n}$.

These last theorems show that it is possible to define multilateral Luenberger indicators which satisfy the additive circularity property, and for which there exist superlative measures which can be calculated without the need for econometric estimation.

6.7 CONCLUSION

This paper has studied the construction of new input and output indicators along the lines suggested by Chambers (1996). Bilateral indicators analogous to those developed in Chambers (1996) have been defined and shown to be calculable under suitable behavioral assumptions directly from observable market data. However, these bilateral Luenberger input and output indicators will only satisfy additive circularity under extreme restrictions on the technology. Consequently, multilateral Luenberger indicators, which satisfy additive circularity and which can be calculated directly from the bilateral indicators, have been defined and shown to be calculable using only data on market prices and quantities.

References

Allais, M. *Traité D'Économie Pure, Vol. 3.* Paris, France: Imprimerie Nationale, 1943.

Bennet, T. L. "The Theory of Measurement of Changes in Cost of Living." *Journal of the Royal Statistical Society* 83 (1920): 445-62.

Blackorby, C., and D. Donaldson. "A Theoretical Treatment of Indices of Absolute Inequality." *International Economic Review* 21, no. 1 (February 1980): 107-36.

Blackorby, C., D. Primont, and R. R. Russell. *Duality, Separability, and Functional Structure.* New York: Elseveir/North-Holland, 1978.

Bowley, A. L. "Notes on Index Numbers." *Economic Journal* 38 (1928): 216-37.

Caves, D. W., L. R. Christensen, and W. E. Diewert. "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity." *Econometrica* 50 (November 1982a): 1393-414.

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

Chambers, R. G. "Consumers' Surplus As an Exact and Superlative Cardinal Measure." Working Paper, University of Maryland, College Park, January 1996a.

Chambers, R. G. "A New Look at Input, Output, Technical Change and Productivity Measurement." Working Paper 96-03, University of Maryland, College Park, January 1996.

Chambers, R. G., Y. Chung, and R. Färe. "Benefit and Distance Functions." *Journal of Economic Theory* 70 (August 1996): 407-19.

Chambers, R. G., Y. Chung, and R. Färe. "Profit, Distance Functions and Nerlovian Efficiency." University of Maryland Working Paper, 1996.

Diewert, W. E. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4 (1976): 115-45.

Diewert, W. E. "Exact and Superlative Welfare Indicators." *Economic Inquiry* 30 (October 1992): 565-82.

Diewert, W. E. "The Measurement of Productivity: A Survey." Paper Presented to Swan Consultants, Ltd. Conference, February 1993.

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

Luenberger, D. G. *Microeconomic Theory.* New York: McGraw Hill, 1995.

Luenberger, D. G. "New Optimality Principles for Economic Efficiency and Equilibrium." *Journal of Optimization Theory and Applications* 75, no. 2 (1992b): 221-64.

Malmquist, S. "Index Numbers and Indifference Surfaces." *Trabajos De Estatistica* 4 (1953): 209-42.

Samuelson, P. A., and S. Swamy. "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis." *American Economic Review* 64 (1974): 566-93.

Appendix

Contributors

Bert Balk
Statistic Methods
Statistics Netherlands
P.O. Box 959
2270 AZ Voorburg
The Netherlands
Fax (31) 703877429

Hans Bjurek
Department of Economics
Göteborgs Universitet
S-41180 Göteborg

Robert Chambers
2200 Symons Hall
Department of Agricultural and Resource Economics
University of Maryland
College Park, MD 20742
USA

Rolf Färe
Department of Economics
Southern Illinois University
Carbondale, IL 62901
USA

Finn Førsund
Department of Economics
University of Oslo
Box 1095
0317 Oslo
NORWAY

Shawna Grosskopf
Department of Economics
Southern Illinois University
Carbondale, IL 62901
USA

Lennart Hjalmarsson
Department of Economics
Göteborgs Universitet
S-41180 Göteborg

Nils-Olov Norlander
Apoteksbolaget
10514 Stockholm
SWEDEN

Pontus Roos
IHE Box 2127
2200 02 Lund
SWEDEN

R. Robert Russell
Department of Economics
University of California, Riverside
Riverside, CA 92521
USA

INDEX