Chapter 1
Overview

This book describes a coherent framework for analysing managerial performance. The focus is on measures of performance that are useful for policy makers. The title of the book reflects the fact that most, if not all, of these measures can be viewed as measures of productivity and/or efficiency. This chapter provides an overview of the main concepts and analytical methods described later in the book.

1.1 Basic Concepts and Terminology

The first step in analysing managerial performance is to identify the manager(s). A manager is a person or other accountable body responsible for controlling (or administering) a firm. In this book, the term 'firm' refers to a production unit (e.g., a school, an assembly line, or an economy). Firm managers are decision makers. For this reason, firms are often[1] referred to as decision-making units (DMUs).

Assessments of managerial performance often depend on the way different variables involved in production processes are classified. In this book, all of the possibly millions of variables that are *physically* involved in production processes are classified into those that are controlled by managers and those that are not. Those that are controlled by managers are then further classified into inputs (i.e., products and services that go *in* to production processes) and outputs (i.e., products and services that come *out* of production processes). Those that are *never* controlled by managers are referred to as environmental variables (e.g., rainfall in crop production). Classifying variables in this way means that managers will not be held responsible for the effects of variables they do not control. For example, farm managers will not be labelled as inefficient when relatively low crop yields are due to low rainfall,

[1] See, for example, Charnes et al. (1981), Cooper et al. (2004) and Färe and Grosskopf (2010).

C. J. O'Donnell, *Productivity and Efficiency Analysis*,
https://doi.org/10.1007/978-981-13-2984-5_1

and truckers will not be labelled as inefficient when delivery delays are due to poor roads. Unless explicitly stated otherwise, the term ‘environmental variable’ is used in this book to refer to a characteristic of a *production* environment. Characteristics of production environments are variables that are *physically* involved in production processes. They should not be confused with characteristics of market environments (e.g., the degree of competition in output markets) or institutional environments (e.g., laws that prevent the use of child labour). Characteristics of market and institutional environments do not generally affect the input-output combinations that are physically possible (i.e., they do not affect the physics). However, as we shall see, they often affect the input-output combinations that managers choose.

One of the most important concepts in efficiency and productivity analysis is the concept of a production technology. In this book, a production technology (or simply ‘technology’) is defined as a technique, method or system for transforming inputs into outputs (e.g., a technique for transforming seeds and other inputs into vegetables). For most practical purposes, it is convenient to think of a technology as a book of instructions, or recipe. The set of technologies that exist in a given period is called a ‘technology set’ (e.g., the set of sustainable, hydroponic, organic, multilayer and vertical-farming techniques for growing vegetables). If we think of a technology as a book of instructions, or recipe, then we can think of a technology set as a library. Measures of ‘technical efficiency’ are viewed as measures of how well technologies are chosen and used (i.e., how well managers ‘choose books/recipes from the library’ and ‘follow the instructions’). The term ‘technical progress’ refers to the discovery of new technologies. Investigative activities aimed at discovering new technologies are referred to as ‘research and development’ (R&D) activities. The term ‘technical regress’ refers to the loss of existing technologies. An important assumption that is maintained throughout the book is that there is no technical regress (i.e., as a society, we do not forget the techniques, methods and systems we know).

The input-output combinations that are possible using different technologies can usually be represented by distance, revenue, cost and/or profit functions. The existence of these functions has few, if any, implications for managerial behaviour. The existence of a cost function, for example, does not imply that managers will aim to minimise costs. Rather, different managers will tend to behave differently depending on what they value, and on what they can and cannot choose. For example, if managers value goods and services at market prices, then, if possible, they will tend to choose inputs and outputs to maximise profits. On the other hand, if managers value products and services differently to the market, then they may instead choose inputs and outputs to maximise measures of productivity. In this book, measures of productivity are defined as measures of output quantity divided by measures of input quantity. Government and community interest in productivity stems from the fact that productivity change is often associated with changes in social welfare; according to Kendrick (1961, p. 3), for example, “[t]he story of productivity, the ratio of output to input, is at heart the record of man’s efforts to raise himself from poverty”.

Decision makers are often interested in measuring levels of efficiency. Measures of efficiency can be viewed *ex post* measures of how well firm managers have solved different optimisation problems. For example, measures of output-oriented technical efficiency can be viewed as measures of how well managers have maximised outputs when inputs and output mixes have been predetermined. On the other hand, measures of profit efficiency can be viewed as measures of how well managers have maximised profits when inputs and outputs have been chosen freely.

Many decision makers are also interested in measuring productivity. This involves assigning numbers to baskets of inputs and outputs. Measurement theory says that so-called index numbers must be assigned in such a way that the relationships between the numbers reflect the relationships between the baskets. For example, if we are measuring changes in output quantities, and if basket A contains exactly twice as much of every output as basket B, then the index number assigned to basket A should be exactly twice as big as the number assigned to basket B. Index numbers that are consistent with measurement theory can be computed using various additive, multiplicative, primal and dual indices (i.e., formulas). Most of the indices currently used in the productivity and efficiency literature yield numbers that are *not* consistent with measurement theory.

Measuring changes in productivity is one thing. Explaining changes in productivity is another. In this book, changes in productivity are explained using a combination of economic theory, measurement theory and statistical methods. Using this so-called econometric approach, changes in productivity can be attributed to four main factors: (a) technical progress (i.e., the discovery of new technologies), (b) environmental change (i.e., changes in variables that are physically involved in production processes but *never* controlled by managers), (c) technical efficiency change (i.e., changes in how well technologies are chosen and used) and (d) scale and mix efficiency change (i.e., changes in economies of scale and substitution). In practice, estimating these different components involves estimating changes in the limits to production (i.e., changes in production frontiers). As we shall see, the choice of estimator depends partly on what is known, or assumed, about production technologies.

1.2 Production Technologies

It is common to make assumptions about technologies by way of assumptions about what they can and cannot produce. For example, it is common to assume that, with a given set of technologies,

- A1 it is possible to produce zero output (i.e., inactivity is possible);
- A2 there is a limit to what can be produced using a finite amount of inputs (i.e., output sets are bounded);
- A3 a positive amount of at least one input is needed in order to produce a strictly positive amount of any output (i.e., inputs are weakly essential; there is 'no free lunch');
- A4 the set of outputs that can be produced using given inputs contains all the points on its boundary (i.e., output sets are closed);

A5 the set of inputs that can produce given outputs contains all the points on its boundary (i.e., input sets are closed);
A6 if particular inputs can be used to produce a given output vector, then they can also be used to produce a scalar contraction of that output vector (i.e., outputs are weakly disposable); and
A7 if particular outputs can be produced using a given input vector, then they can also be produced using a scalar magnification of that input vector (i.e., inputs are weakly disposable).

Assumptions A1–A7 are maintained throughout this book. Other assumptions that are made from time to time include the following:

A6s if given inputs can be used to produce particular outputs, then they can also be used to produce fewer outputs (i.e., outputs are strongly disposable);
A7s if given outputs can be produced using particular inputs, then they can also be produced using more inputs (i.e., inputs are strongly disposable);
A8s if a given output-input combination is possible in a particular production environment, then it is also possible in a better production environment (i.e., environmental variables are strongly disposable).

The word 'strong' is used in A6s and A7s to reflect the fact that A6s implies A6 and A7s implies A7 (symbolically, A6s $\Rightarrow$ A6 and A7s $\Rightarrow$ A7). The input-output combinations that are possible using different sets of technologies can be represented by output sets, input sets and production possibilities sets. If A2, A6 and A7 are true, then they can also be represented by output and input distance functions.

1.2.1 Output Sets

An output set is a set containing all outputs that can be produced using given inputs. In this book, the focus is on *period-and-environment-specific* output sets. A period-and-environment-specific output set is a set containing all outputs that can be produced using given inputs *in a given period in a given production environment*. For a precise definition, let $x = (x_1, \ldots, x_M)'$, $q = (q_1, \ldots, q_N)'$ and $z = (z_1, \ldots, z_J)'$ denote vectors of nonnegative inputs, outputs and environmental variables (respectively). In mathematical terms, the set of outputs that can be produced using the input vector x in period t in a production environment characterised by z is

$$P^t(x, z) = \{q : x \text{ can produce } q \text{ in period } t \text{ in environment } z\}. \quad (1.1)$$

To illustrate, Table 1.1 reports artificial (or 'toy') data on $I = 5$ firms over $T = 5$ time periods. Each firm has used two inputs to produce two outputs in one of two production environments. Figure 1.1 depicts the set of outputs that could have been produced using the input vector $\iota = (1, 1)'$ in period 1 in environment 1. The dots in this figure mark the observed output combinations of the two firms that used this

Table 1.1 Toy data

Row	Firm	Period	q_1	q_2	x_1	x_2	z
A	1	1	1	1	1	1	1
B	2	1	1	1	0.56	0.56	1
C	3	1	2.37	2.37	1	1	1
D	4	1	2.11	2.11	1.05	0.7	1
E	5	1	1.81	3.62	1.05	0.7	1
F	1	2	1	1	0.996	0.316	2
G	2	2	1.777	3.503	1.472	0.546	2
H	3	2	0.96	0.94	0.017	0.346	1
I	4	2	5.82	0.001	4.545	0.01	2
J	5	2	6.685	0.001	4.45	0.001	1
K	1	3	1.381	4.732	1	1	1
L	2	3	0.566	4.818	1	1	1
M	3	3	1	3	1.354	1	1
N	4	3	0.7	0.7	0.33	0.16	1
O	5	3	2	2	1	1	2
P	1	4	1	1	0.657	0.479	1
R	2	4	1	3	1	1	1
S	3	4	1	1	1.933	0.283	2
T	4	4	1.925	3.722	1	1	2
U	5	4	1	1	1	0.31	1
V	1	5	1	5.166	1	1	1
W	2	5	2	2	0.919	0.919	2
X	3	5	1	1	1.464	0.215	2
Y	4	5	1	1	0.74	0.74	1
Z	5	5	1.81	3.62	2.1	1.4	1

input vector in this period in this environment (in this book, letters in figures generally correspond to rows in tables). The set $P^1(\iota, 1)$ is the area bounded by the two axes and the curve passing through point C.

1.2.2 *Input Sets*

An input set is a set containing all inputs that can produce given outputs. Again, this book focuses on *period-and-environment-specific* input sets. A period-and-environment-specific input set is a set containing all inputs that can produce given outputs *in a given period in a given production environment*. For example, the set of inputs that can produce the output vector q in period t in an environment characterised

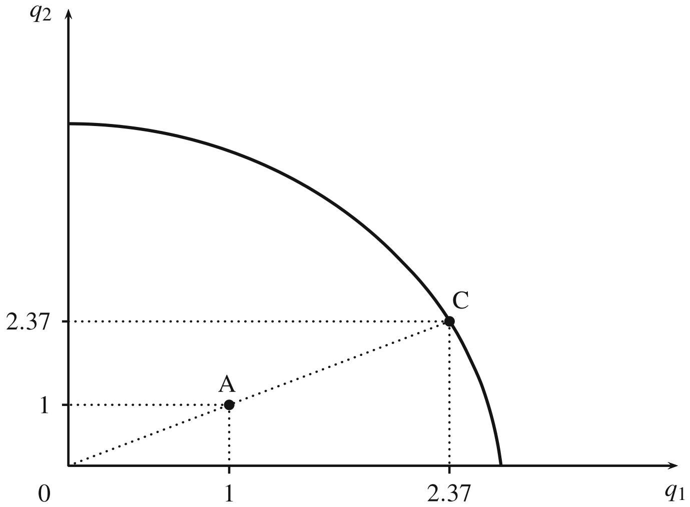

Fig. 1.1 The outputs that could have been produced using one unit of each input in period 1 in environment 1. The set $P^1(\iota, 1)$ is the area bounded by the two axes and the curve passing through point C

by z is

$$L^t(q, z) = \{x : x \text{ can produce } q \text{ in period } t \text{ in environment } z\}. \quad (1.2)$$

To illustrate, reconsider the toy data in Table 1.1. Figure 1.2 depicts the set of inputs that could have produced one unit of each output in period 1 in environment 1. The dots in this figure mark the observed input combinations of the two firms that produced these outputs in this period in this environment. The set $L^1(\iota, 1)$ comprises all points on and above the curve passing through point B.

1.2.3 Production Possibilities Sets

A production possibilities set is a set containing all input-output combinations that are physically possible. In this book, the focus is on two specific types of production possibilities set: *period-and-environment-specific* production possibilities sets and period-environment-*and-mix-specific* production possibilities sets.

A *period-and-environment-specific* production possibilities set is a set containing all input-output combinations that are physically possible *in a given period in a given production environment*. For example, the set of input-output combinations that are physically possible in period t in a production environment characterised by z is

$$T^t(z) = \{(x, q) : x \text{ can produce } q \text{ in period } t \text{ in environment } z\}. \quad (1.3)$$

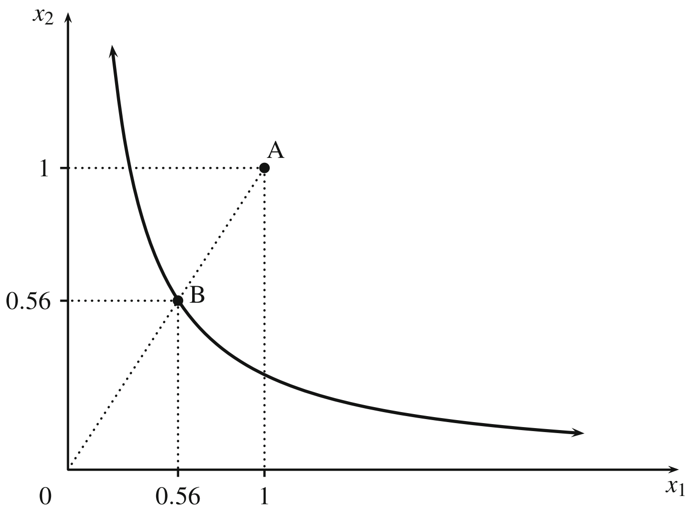

Fig. 1.2 The inputs that could have produced one unit of each output in period 1 in environment 1. The set $L^1(\iota, 1)$ comprises all points on and above the frontier passing through point B

If there are more than two outputs and inputs, then the only way to represent this set in a two-dimensional figure is to map many variables into just two variables. Throughout this book, outputs are mapped into scalar-valued measures of total output and inputs are mapped into scalar-valued measures of total input. In the case of outputs, the measure of total (or aggregate) output associated with the vector q is given by $Q(q)$, where $Q(.)$ is any nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator function. In the case of inputs, the measure of total (or aggregate) input associated with the vector x is given by $X(x)$, where, again, $X(.)$ is any nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator function.

For a simple illustration, reconsider the toy data in Table 1.1, and let $Q(q) = 0.484q_1 + 0.516q_2$ and $X(x) = 0.23x_1 + 0.77x_2$. The associated aggregate outputs and inputs are reported in Table 1.2. Figure 1.3 plots the aggregate outputs and inputs of the five firms that operated in period 1 in environment 1. In this figure, the set $T^1(1)$ is represented by the area bounded by the horizontal axis and the curve passing through point E.

A period-environment-*and-mix-specific* production possibilities set is a set containing all input-output combinations that are physically possible *when using a scalar multiple of a given input vector to produce a scalar multiple of a given output vector* in a given period in a given production environment. For example, the set of input-output combinations that are possible when using a scalar multiple of $\bar{x}$ to produce a scalar multiple of $\bar{q}$ in period t in an environment characterised by z is

$$T^t(\bar{x}, \bar{q}, z) = \{(x, q) : x \propto \bar{x}, \ q \propto \bar{q}, \ (x, q) \in T^t(z)\}. \quad (1.4)$$

Table 1.2 Aggregate outputs and inputs[a]

Row	Firm	Period	$Q(q)$	$X(x)$
A	1	1	1	1
B	2	1	1	0.56
C	3	1	2.37	1
D	4	1	2.11	0.7805
E	5	1	2.744	0.7805
F	1	2	1	0.472
G	2	2	2.668	0.759
H	3	2	0.950	0.270
I	4	2	2.817	1.053
J	5	2	3.236	1.024
K	1	3	3.110	1
L	2	3	2.76	1
M	3	3	2.032	1.081
N	4	3	0.7	0.199
O	5	3	2	1
P	1	4	1	0.520
R	2	4	2.032	1
S	3	4	1	0.663
T	4	4	2.852	1
U	5	4	1	0.469
V	1	5	3.150	1
W	2	5	2	0.919
X	3	5	1	0.502
Y	4	5	1	0.74
Z	5	5	2.744	1.561

[a]Numbers reported to less than three decimal places are exact in the sense that they have not been rounded. Some of the other numbers may have been rounded

To illustrate, reconsider the toy data in Tables 1.1 and 1.2. Figure 1.4 plots the aggregate outputs and inputs of the three firms that used a scalar multiple of ι to produce a scalar multiple of ι in period 1 in environment 1. In this figure, the set $T^1(\iota, \iota, 1)$ is represented by the area bounded by the horizontal axis and the curve passing through points B and C.

1.2.4 Output Distance Functions

Set representations of technologies can be difficult to work with mathematically. In practice, it is common to work with distance functions. An output distance function

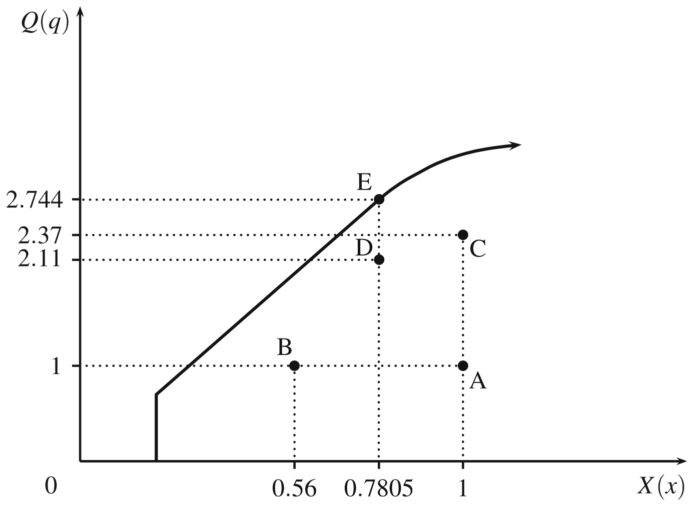

Fig. 1.3 The input-output combinations that were possible in period 1 in environment 1. The set $T^1(1)$ is the area bounded by the horizontal axis and the curve passing through point E

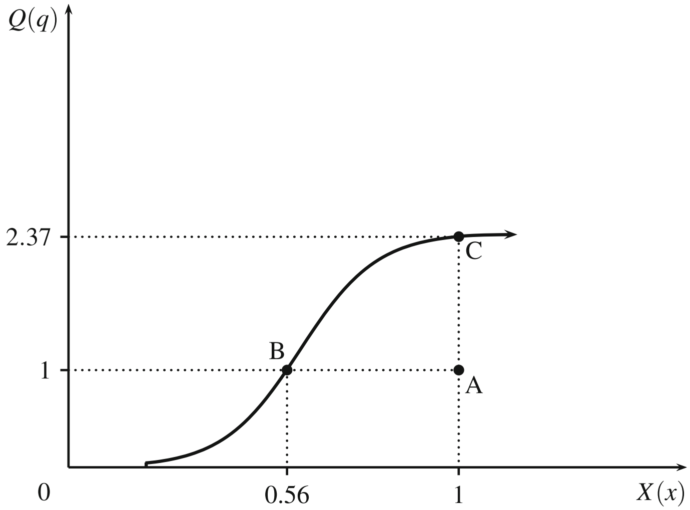

Fig. 1.4 The input-output combinations that were possible when using a scalar multiple of ι to produce a scalar multiple of ι in period 1 in environment 1. The set $T^1(\iota, \iota, 1)$ is the area bounded by the horizontal axis and the curve passing through points B and C

gives the reciprocal of the largest factor by which it is possible to scale up a given output vector when using a given input vector. For example, if it is technically possible to use given inputs to produce four times as much of every output, then the output distance function takes the value $1/4 = 0.25$. Again, this book focuses on *period-and-environment-specific* output distance functions. A period-and-environment-specific output distance function gives the reciprocal of the largest factor by which it is possible to scale up a given output vector when using a given input vector *in a given period in a given production environment*. For example, the reciprocal of the largest factor by which it is possible to scale up q when using x in period t in environment z is

$$D_O^t(x, q, z) = \inf\{\rho > 0 : q/\rho \in P^t(x, z)\}. \tag{1.5}$$

For a numerical example, reconsider the toy data reported in Table 1.1. The outputs of firm 1 in period 1 (hereafter firm A) were previously mapped to point A in Fig. 1.1. That figure reveals that it would have been technically possible to hold the inputs of the firm fixed and scale up its outputs by a factor of no more than 2.37. Thus, in the case of firm A, the output distance function takes the value $D_O^1(\iota, \iota, 1) = 1/2.37 = 0.422$.

1.2.5 Input Distance Functions

An input distance function gives the reciprocal of the smallest fraction of a given input vector that can produce a given output vector. For example, if it is technically possible to produce a given output vector using as little as one-half of a given input vector, then the input distance function takes the value $1/0.5 = 2$. Again, this book focuses on *period-and-environment-specific* input distance functions. A period-and-environment-specific input distance function gives the reciprocal of the smallest fraction of a given input vector that can produce a given output vector *in a given period in a given production environment*. For example, the reciprocal of the smallest fraction of x that can produce q in period t in environment z is

$$D_I^t(x, q, z) = \sup\{\theta > 0 : x/\theta \in L^t(q, z)\}. \tag{1.6}$$

For a numerical example, reconsider the toy data in Table 1.1. The inputs of firm 1 in period 1 (i.e., firm A) were previously mapped to point A in Fig. 1.2. That figure reveals that it would have been technically possible to produce the outputs of the firm using as little as $0.56/1 = 56\%$ of its inputs. Thus, in the case of firm A, the input distance function takes the value $D_I^1(\iota, \iota, 1) = 1/0.56 = 1.786$.

1.2.6 Other Sets and Functions

If assumptions A1–A7 are true, then the input-output combinations that are possible using different technologies can also be represented by revenue and cost functions. A revenue function gives the maximum revenue that can be earned using given inputs. A cost function gives the minimum cost of producing given outputs. Other sets and functions that are discussed in this book include profit functions, production functions, input requirement functions, directional distance functions, hyperbolic distance functions, technology-and-environment-specific sets and functions, period-specific sets and functions, and state-contingent sets and functions.

1.3 Measures of Productivity Change

In this book, measures of productivity change are defined as measures of output quantity change divided by measures of input quantity change. Computing measures of output and input quantity change involves assigning numbers to baskets of outputs and inputs. Measurement theory says that so-called index numbers cannot be assigned in an arbitrary way. Rather, they must be assigned in such a way that the relationships between the numbers mirror the relationships between the baskets. To illustrate, consider the baskets of maple syrup and Vegemite and the associated sets of quantity index numbers presented in Table 1.3. Among other things, the index numbers in column L indicate that basket W contains twice as much syrup and Vegemite as basket A. The other index numbers in the table can be interpreted in a similar way. The index numbers in column L are the only numbers that are consistent with measurement

Table 1.3 Quantity index numbers

		L	F	CF	EKS
Basket A		1	1	1	1
Basket M		2.032	1.892	2.389	1.942
Basket R		2.032	1.893	2.854	1.943
Basket W		2	2	3.642	2.027

theory. Observe, for example, that basket M contains the same amount of syrup and Vegemite as basket R, and only in column L is the index number in row M the same as the index number in row R. Arguably the most important distinguishing feature of this book is that it assigns numbers to baskets of outputs and inputs in a way that is consistent with measurement theory. To clarify the approach, this section introduces firm and time subscripts into the notation. Thus, for example, $q_{it} = (q_{1it}, \ldots, q_{Nit})'$ and $x_{it} = (x_{1it}, \ldots, x_{Mit})'$ now denote the output and input vectors of firm i in period t.

1.3.1 Output Quantity Indices

An *index* is a rule or a formula that tells us how to use data to measure the change in one or more variables over time and/or space. An *index number* is the value obtained after data have been substituted into the formula. In this book, an output quantity index (or simply 'output index') that compares q_{it} with q_{ks} is defined as any variable of the form

$$QI(q_{ks}, q_{it}) \equiv Q(q_{it})/Q(q_{ks}) \tag{1.7}$$

where $Q(.)$ is a nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator function. All output indices of this type yield numbers that are consistent with measurement theory. They are also proper indices in the sense that, if outputs are positive, then they satisfy axioms Q1 to Q8 listed in O'Donnell (2016). Two of the most important axioms are a transitivity axiom and a proportionality axiom. The transitivity axiom says that a direct comparison of the outputs of two firms should yield the same index number as an indirect comparison through a third firm. If, for example, firm R produced the same amount of every output as firm M, and firm M produced λ times as much as firm A, then the index that compares the outputs of firm R with the outputs of firm A must take the value λ (indicating that firm R produced λ times as much as firm A). The proportionality axiom says that if firm W produced λ times as much as firm A, then the index that compares the outputs of firm W with the outputs of firm A must take the value λ. The class of proper output indices includes the Lowe index defined by O'Donnell (2012, Eq. 3). Output indices that do not satisfy the transitivity axiom and are therefore not proper include the well-known Fisher and Törnqvist indices. Output indices that do not satisfy the proportionality axiom and are therefore not proper include the chained Fisher (CF) index and an index proposed by Elteto and Koves (1964) and Szulc (1964) (hereafter, EKS).

To illustrate, consider the output quantities, output prices and output index numbers reported in Table 1.4. The output quantities reported in this table are the quantities reported earlier in Table 1.1. The index numbers in the different columns are Lowe (L), Fisher (F), CF and EKS index numbers that compare the output quantities in each row with the output quantities in row A. The Lowe index numbers were computed using the same aggregator function that was used to compute the aggregate

Table 1.4 Output quantities, output prices and output index numbers[a,b]

Row	q_1	q_2	p_1	p_2	L	F	CF	EKS
A	1	1	0.57	0.41	1	1	1	1
B	1	1	0.26	0.25	1	1	1	0.992*
C	2.37	2.37	0.57	0.41	2.37	2.37	2.37	2.37
D	2.11	2.11	0.58	0.53	2.11	2.11	2.11	2.096*
E	1.81	3.62	0.26	0.26	2.744	2.640*	2.695*	2.677*
F	1	1	0.59	0.76	1	1	0.972*	0.986*
G	1.777	3.503	0.63	0.65	2.668	2.575	2.626	2.608
H	0.96	0.94	0.34	0.31	0.950	0.951	0.950	0.944
I	5.82	0.001	0.46	0.58	2.817	2.952	2.800	2.672
J	6.685	0.001	0.61	1.43	3.236	2.789	3.217	2.508
K	1.381	4.732	0.57	0.41	3.110	2.783	3.716	2.883
L	0.566	4.818	0.49	0.65	2.760	2.648	3.251	2.737
M	1	3	0.51	0.46	2.032	1.892*	2.389*	1.942*
N	0.7	0.7	0.52	0.23	0.7	0.7	0.943*	0.711*
O	2	2	0.37	0.17	2	2	2.695*	2.029*
P	1	1	0.41	0.76	1	1	1.348*	0.982*
R	1	3	0.53	0.48	2.032	1.893*	2.854*	1.943*
S	1	1	0.53	0.37	1	1	1.514*	1.001*
T	1.925	3.722	0.91	0.53	2.852	2.631	3.973	2.706
U	1	1	0.31	1.03	1	1	1.359*	0.981*
V	1	5.166	0.47	0.08	3.150	2.099	3.530	2.296
W	2	2	0.57	0.27	2	2	3.642*	2.027*
X	1	1	0.31	0.51	1	1	1.821*	0.983*
Y	1	1	0.31	0.67	1	1	1.821*	0.981*
Z	1.81	3.62	0.42	0.69	2.744	2.745*	5.447*	2.759*

[a]L = Lowe; F = Fisher; CF = chained Fisher; EKS = Elteto-Koves-Szulc
[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8
*Incoherent (not because of rounding)

outputs in Table 1.2. Lowe index numbers are consistent with measurement theory. Observe, for example, that the output vector in row M is the same as the output vector in row R, and the Lowe index number in row M is the same as the Lowe index number in row R (the index numbers in these particular rows are, in fact, the index numbers reported above in Table 1.3). The Fisher, CF and EKS index numbers are *not* consistent with measurement theory.[2] Numbers that are clearly incoherent are marked with an asterisk (*). Observe, for example, that the outputs in row E are the

[2]In practice, CF (resp. EKS) indices are mainly used for time-series (resp. cross-section) comparisons. For this reason, the CF numbers in Table 1.4 were computed by treating the observations in the dataset as observations on one firm over twenty-five periods. The EKS numbers were computed by treating the observations in the dataset as observations on twenty-five firms in one period.

same as the outputs in row Z, but the CF index number in row E differs from the CF index number in row Z.

1.3.2 Input Quantity Indices

In this book, an input quantity index (or simply 'input index') that compares x_{it} with x_{ks} is defined as any variable of the form

$$XI(x_{ks}, x_{it}) \equiv X(x_{it})/X(x_{ks}) \tag{1.8}$$

where $X(.)$ is a nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator function. Again, all input indices of this type yield numbers that are consistent with measurement theory. They are also proper indices in the sense that, if inputs are positive, then they satisfy axioms X1 to X8 listed in O'Donnell (2016). The class of proper input indices includes the Lowe index defined by O'Donnell (2012, Eq. 4). Again, input indices that are not proper include the Fisher, CF and EKS indices.

To illustrate, consider the input quantities, input prices and input index numbers reported in Table 1.5. The input quantities reported in this table are the quantities reported earlier in Table 1.1. The index numbers in the different columns are Lowe (L), Fisher (F), CF and EKS index numbers that compare the input quantities in each row with the input quantities in row A. The Lowe index numbers were computed using the same aggregator function that was used to compute the aggregate inputs in Table 1.2. Again, these numbers are consistent with measurement theory. For example, the input vector in row D is the same as the input vector in row E, and the Lowe index number in row D is the same as the Lowe index number in row E. Again, the Fisher, CF and EKS index numbers are *not* consistent with measurement theory.[3] Again, numbers that are clearly incoherent are marked with an asterisk (*). Observe, for example, that the input vector in row Z is twice as big as the input vector in row E, but the EKS index number in row Z is not twice as big as the EKS index number in row E.

1.3.3 Productivity Indices

Productivity indices are measures of productivity change. Without loss of generality, this book focuses on measures of total factor productivity (TFP) change. An index that compares the TFP of firm i in period t with the TFP of firm k in period s is defined

[3] Again, the CF numbers were computed by treating the observations in the dataset as observations on one firm over twenty-five periods. The EKS index numbers were computed by treating the observations in the dataset as observations on twenty-five firms in one period.

Table 1.5 Input quantities, input prices and input index numbers[a,b]

Row	x_1	x_2	w_1	w_2	L	F	CF	EKS
A	1	1	0.28	1.91	1	1	1	1
B	0.56	0.56	0.22	0.58	0.56	0.56	0.56	0.525*
C	1	1	0.28	1.91	1	1	1	1
D	1.05	0.7	0.16	0.41	0.781	0.771*	0.771*	0.749*
E	1.05	0.7	0.07	1.02	0.781	0.734*	0.771*	0.797*
F	0.996	0.316	0.24	0.29	0.472	0.501	0.464	0.502
G	1.472	0.546	0.16	0.16	0.759	0.819	0.715	0.798
H	0.017	0.346	0.17	0.7	0.270	0.293	0.189	0.253
I	4.545	0.01	0.27	0.39	1.053	1.049	1.001	1.339
J	4.45	0.001	0.29	0.79	1.024	0.825	0.976	1.102
K	1	1	0.28	1.91	1	1	1.182*	1
L	1	1	0.21	0.56	1	1	1.182*	0.939*
M	1.354	1	0.16	0.74	1.081	1.054	1.276	1.056
N	0.33	0.16	0.24	2.3	0.199	0.179	0.223	0.196
O	1	1	0.24	0.15	1	1	1.032*	0.863*
P	0.657	0.479	0.26	0.61	0.520	0.517	0.578	0.495
R	1	1	0.16	0.22	1	1	1.064*	0.899*
S	1.933	0.283	0.19	0.62	0.663	0.575	0.861	0.668
T	1	1	0.17	0.26	1	1	1.088*	0.905*
U	1	0.31	0.27	0.91	0.469	0.432	0.568	0.464
V	1	1	0.29	0.78	1	1	1.178*	0.939*
W	0.919	0.919	0.39	0.81	0.919	0.919	1.083*	0.848*
X	1.464	0.215	0.21	0.31	0.502	0.519	0.787	0.572
Y	0.74	0.74	0.23	0.69	0.74	0.74	0.946*	0.700*
Z	2.1	1.4	0.31	0.22	1.561	1.642*	2.159*	1.479*

[a]L = Lowe; F = Fisher; CF = chained Fisher; EKS = Elteto-Koves-Szulc
[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8
*Incoherent (not because of rounding)

as any variable of the form $TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) \equiv QI(q_{ks}, q_{it})/XI(x_{ks}, x_{it})$ where $QI(.)$ is any proper output index and $XI(.)$ is any proper input index. Equivalently,

$$TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) \equiv TFP(x_{it}, q_{it})/TFP(x_{ks}, q_{ks}) \tag{1.9}$$

where $TFP(x_{it}, q_{it}) \equiv Q(q_{it})/X(x_{it})$ denotes the TFP of firm i in period t. All TFP indices (TFPIs) of this type are said to be proper. If outputs and inputs are positive, then they satisfy axioms T1 to T8 listed in O'Donnell (2017). The class of proper TFPIs includes the Lowe index defined by O'Donnell (2012, Eq. 5). TFPIs that are not proper include the Fisher, CF and EKS indices.

Table 1.6 Output quantities, input quantities and TFPI numbers[a,b]

Row	q_1	q_2	x_1	x_2	L	F	CF	EKS
A	1	1	1	1	1	1	1	1
B	1	1	0.56	0.56	1.786	1.786	1.786	1.889*
C	2.37	2.37	1	1	2.37	2.37	2.37	2.37
D	2.11	2.11	1.05	0.7	2.703	2.737	2.737	2.799
E	1.81	3.62	1.05	0.7	3.516	3.599*	3.495*	3.359*
F	1	1	0.996	0.316	2.117	1.994	2.096	1.963
G	1.777	3.503	1.472	0.546	3.515	3.145	3.670	3.269
H	0.96	0.94	0.017	0.346	3.513	3.250	5.028	3.728
I	5.82	0.001	4.545	0.01	2.675	2.815	2.798	1.996
J	6.685	0.001	4.45	0.001	3.159	3.378	3.296	2.276
K	1.381	4.732	1	1	3.110	2.783	3.144	2.883
L	0.566	4.818	1	1	2.760	2.648	2.750	2.916
M	1	3	1.354	1	1.879	1.795	1.872	1.840
N	0.7	0.7	0.33	0.16	3.516	3.913	4.233	3.629
O	2	2	1	1	2	2	2.611*	2.350*
P	1	1	0.657	0.479	1.923	1.935	2.332	1.985
R	1	3	1	1	2.032	1.893	2.682	2.162
S	1	1	1.933	0.283	1.509	1.738	1.757	1.498
T	1.925	3.722	1	1	2.852	2.631	3.652	2.991
U	1	1	1	0.31	2.134	2.317	2.391	2.117
V	1	5.166	1	1	3.150	2.099	2.996	2.445
W	2	2	0.919	0.919	2.176	2.176	3.364*	2.390*
X	1	1	1.464	0.215	1.991	1.926	2.313	1.719
Y	1	1	0.74	0.74	1.351	1.351	1.925*	1.401*
Z	1.81	3.62	2.1	1.4	1.758	1.672*	2.523*	1.866*

[a] L = Lowe; F = Fisher; CF = chained Fisher; EKS = Elteto-Koves-Szulc
[b] Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8
*Incoherent (not because of rounding)

To illustrate, consider the output quantities, input quantities and TFPI numbers reported in Table 1.6. The output and input quantities reported in this table are the quantities reported earlier in Tables 1.1, 1.4 and 1.5. The TFPI numbers in the columns labelled L, F, CF and EKS were obtained by dividing the output index numbers in Table 1.4 by the corresponding input index numbers in Table 1.5. The index numbers in Table 1.6 compare the output-input combinations in each row with the output-input combinations in row A. The Lowe index numbers reported in column L are coherent. Observe, for example, that (a) the output vector in row W is twice as big as the output vector in row A, (b) the input vector in row W is only 0.919 times as big as the input vector in row A, and (c) the Lowe TFPI number is $2/0.919 = 2.176$. Again, the Fisher, CF and EKS index numbers are *not* coherent. Observe, for example, that (a) the output vector in row Z is the same as the output vector in row E, (b) the input vector in row Z is twice as big as the input vector in row E, but (c) the CF index number in row Z is not half as big as the CF index number in row E.

1.3.4 Other Indices

Other indices discussed in this book include output price indices, input price indices, terms-of-trade (TT) indices, implicit output indices, implicit input indices and implicit productivity indices. Implicit output (resp. input) indices are obtained by dividing revenue (resp. cost) indices by output price (resp. input price) indices. Implicit productivity indices are obtained by dividing profitability indices by TT indices. Except in restrictive special cases, implicit indices yield numbers that are not consistent with measurement theory.

1.4 Managerial Behaviour

To explain changes in outputs and inputs, and therefore changes in productivity, we need to know something about managerial behaviour. The existence of different sets and functions has few, if any, implications for behaviour. The existence of revenue functions, for example, does not mean that managers will choose outputs in order to maximise revenues, and the existence of cost functions does not mean they will choose inputs to minimise costs. Instead, different managers will tend to behave differently depending on what they value, and on what they can and cannot choose. Some of the simplest optimisation problems faced by firm managers involve maximising outputs, minimising inputs and/or maximising productivity.

1.4.1 Output Maximisation

The managers of some firms (e.g., the managers of government departments, benevolent societies, conservation groups and socially-responsible corporations) often value products and services differently to the market. There are also many products and services that are not exchanged in a market and therefore do not have a market price (e.g., city parks). If a firm manager places nonnegative values on outputs (not necessarily market values) and all other variables involved in the production process have been predetermined (i.e., have been determined in a previous period), then (s)he will generally aim to maximise a measure of total output. If there is more than one output, then the precise form of the output maximisation problem will depend on how easily the manager can choose the output mix. Suppose, for example, the manager of firm i can only choose output vectors that are scalar multiples of q_{it}. In this case, his/her period-t output-maximisation problem can be written as

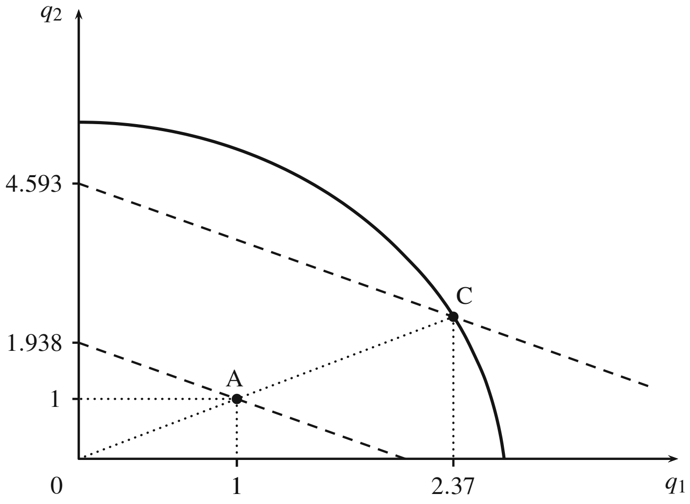

Fig. 1.5 Output maximisation. If the output mix of firm A had been predetermined, then the manager could have maximised total output by operating the firm at point C

$$\max_q \{Q(q) : q \propto q_{it},\ D_O^t(x_{it}, q, z_{it}) \leq 1\} \tag{1.10}$$

where $Q(.)$ is any nonnegative, nondecreasing, linearly-homogeneous, scalar-valued function satisfying $Q(q_{it}) > 0$. The output vector that solves this problem is $\bar{q}_{it} \equiv \bar{q}^t(x_{it}, q_{it}, z_{it}) = q_{it}/D_O^t(x_{it}, q_{it}, z_{it})$. The associated aggregate output is $Q(\bar{q}_{it}) = Q(q_{it})/D_O^t(x_{it}, q_{it}, z_{it})$.

For a numerical example, reconsider the toy data in Table 1.1. Also let $Q(q) = 0.484q_1 + 0.516q_2$. Figure 1.5 depicts the output maximisation problem that would have faced the manager of firm 1 in period 1 (i.e., firm A) had the firm's output mix been predetermined. In this figure, the frontier passing through point C is the frontier depicted earlier in Fig. 1.1. The outputs of firm 1 in period 1 map to point A. The aggregate output at this point is $Q(q_{11}) = 1$. The dashed line passing through point A is an iso-output line with a slope of -0.938 and a q_2 intercept of $Q(q_{11})/0.516 = 1.938$. The other dashed line is an iso-output line with the same slope but a higher intercept. Output maximisation involves choosing the iso-output line with the highest intercept that passes through a technically-feasible point. If the output mix of firm A had been predetermined, then the output-maximising iso-output line would have been the one passing through point C. The aggregate output at this point is $Q(\bar{q}_{11}) = 4.593 \times 0.516 = 2.37$.

1.4.2 Input Minimisation

If a firm manager places nonnegative values on inputs (again, not necessarily market values) and all other variables involved in the production process have been predetermined, then (s)he will generally aim to minimise a measure of total input. If there is more than one input, then the precise form of the input minimisation problem will depend on how easily the manager can choose the input mix. Suppose, for example, the manager of firm i can only use input vectors that are scalar multiples of x_{it}. In this case, his/her period-t input-minimisation problem can be written as

$$\min_{x}\{X(x) : x \propto x_{it},\ D_I^t(x, q_{it}, z_{it}) \geq 1\} \tag{1.11}$$

where $X(.)$ is any nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator function satisfying $X(x_{it}) > 0$. The input vector that solves this problem is $\bar{x}_{it} \equiv \bar{x}^t(x_{it}, q_{it}, z_{it}) = x_{it}/D_I^t(x_{it}, q_{it}, z_{it})$. The associated aggregate input is $X(\bar{x}_{it}) = X(x_{it})/D_I^t(x_{it}, q_{it}, z_{it})$.

For a numerical example, reconsider the toy data in Table 1.1. Also let $X(x) = 0.23x_1 + 0.77x_2$. Figure 1.6 depicts the input minimisation problem that would have faced the manager of firm 1 in period 1 (i.e., firm A) had the firm's input mix been predetermined. In this figure, the frontier passing through point B is the frontier depicted earlier in Fig. 1.2. The inputs of firm 1 in period 1 map to point A. The aggregate input at this point is $X(x_{11}) = 1$. The dashed line passing through point A is an iso-input line with a slope of -0.299 and an x_2 intercept of $X(x_{11})/0.77 = 1.299$. The other dashed line is an iso-input line with the same slope but a lower intercept.

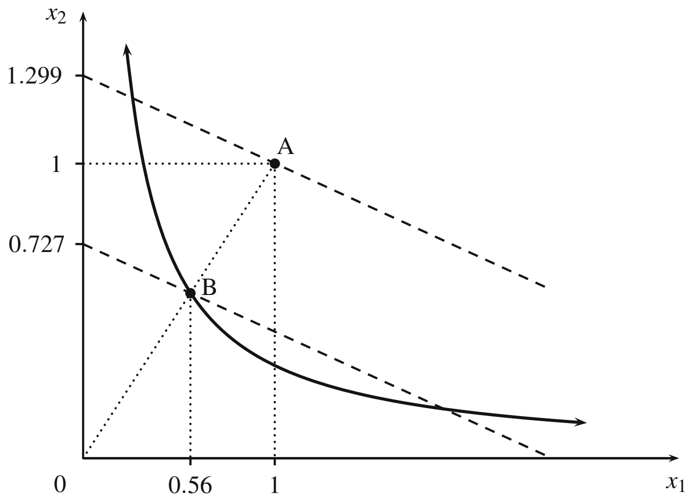

Fig. 1.6 Input minimisation. If the input mix of firm A had been predetermined, then the manager could have minimised total input use by operating the firm at point B

Input minimisation involves choosing the iso-input line with the lowest intercept that passes through a technically-feasible point. If the input mix of firm A had been predetermined, then the input-minimising iso-input line would have been the one passing through point B. The aggregate input at this point is $X(\bar{x}_{11}) = 0.727 \times 0.77 = 0.56$.

1.4.3 Productivity Maximisation

If a firm manager places nonnegative values on outputs and inputs (again, not necessarily market values) and all environmental variables have been predetermined, then (s)he may aim to maximise a measure of TFP. If there is more than one output and more than one input, then the precise form of the manager's TFP maximisation problem will depend on how easily (s)he can choose the output mix and the input mix. Suppose, for example, the manager of firm i can choose all outputs and inputs freely. In this case, his/her period-t TFP-maximisation problem can be written as

$$\max_{q,x}\{Q(q)/X(x) : D^t_O(x, q, z_{it}) \leq 1\} \tag{1.12}$$

where $Q(.)$ and $X(.)$ are nonnegative, nondecreasing, linearly-homogeneous, scalar-valued aggregator functions with parameters (or weights) that represent the values the manager places on outputs and inputs. There may be several pairs of output and input vectors that solve this problem. Let $q^*_{it} \equiv q^t(z_{it})$ and $x^*_{it} \equiv x^t(z_{it})$ denote one such pair. The associated maximum TFP is $TFP^t(z_{it}) = Q(q^*_{it})/X(x^*_{it})$.

For a numerical example, reconsider the toy data in Tables 1.1 and 1.2. Figure 1.7 depicts the TFP maximisation problem that would have faced the manager of firm 1 in period 1 (i.e., firm A). The frontier in this figure is the frontier depicted earlier in Fig. 1.3. The outputs and inputs of firm 1 in period 1 map to point A. The dashed line passing through point A is an iso-productivity ray with a slope of $TFP(x_{11}, q_{11}) =$ slope $0A = 1/1 = 1$. The other dashed lines are iso-productivity rays with higher slopes. TFP maximisation involves choosing the iso-productivity ray that has the highest slope and passes through a technically-feasible point. If the manager of firm A had been able to choose all outputs and inputs freely, then the TFP-maximising iso-productivity ray would have been the one passing through points N and E. The TFP at any point on the line connecting these two points is $TFP^1(z_{11}) =$ slope $0N =$ slope $0E = 0.7/0.1991 = 2.744/0.7805 = 3.516$.

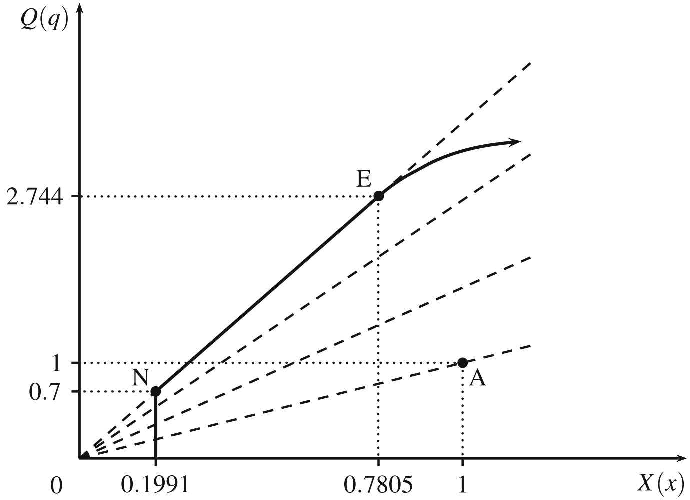

Fig. 1.7 Productivity maximisation. If the manager of firm A had been able to choose all outputs and inputs freely, then (s)he could have maximised TFP by operating the firm anywhere on the line connecting points N and E

1.4.4 Other Types of Behaviour

Other optimisation problems (and therefore other types of managerial behaviour) discussed in this book involve maximising revenue, minimising cost, maximising profit, maximising net output, and maximising return to the dollar.

1.5 Measures of Efficiency

Measures of efficiency can be viewed as *ex post* measures of how well firm managers have solved different optimisation problems. Except where explicitly stated otherwise, all measures of efficiency defined in this book take values in the closed unit interval. A firm manager is said to have been fully efficient by some measure if and only if that measure takes the value one.

1.5.1 Output-Oriented Measures

Output-oriented measures of efficiency are relevant measures of managerial performance in situations where managers have placed nonnegative values on outputs (not necessarily market values) and inputs have been predetermined. In these situations, the relevance of a particular measure depends on how easily the manager has been able to choose the output mix. If, for example, the output mix of the firm has been predetermined, then the most relevant measure is output-oriented technical efficiency (OTE). Several measures of OTE can be found in the literature. In this book, the OTE of manager i in period t is defined as $OTE^t(x_{it}, q_{it}, z_{it}) = D_O^t(x_{it}, q_{it}, z_{it})$. Equivalently,

$$OTE^t(x_{it}, q_{it}, z_{it}) = Q(q_{it})/Q(\bar{q}_{it}) \quad (1.13)$$

where $Q(q_{it})$ is the aggregate output of the firm and $Q(\bar{q}_{it}) = Q(q_{it})/D_O^t(x_{it}, q_{it}, z_{it})$ is the maximum aggregate output that is possible in period t when using x_{it} to produce a scalar multiple of q_{it} in an environment characterised by z_{it}. The right-hand side of (1.13) is, in fact, an output index. If environmental variables have been predetermined, then it can be viewed as a measure of how well the manager has solved problem (1.10).

For a numerical example, reconsider the output maximisation problem depicted earlier in Fig. 1.5. In that figure, the outputs of firm 1 in period 1 were represented by point A. The aggregate output at point A is $Q(q_{11}) = 1.938 \times 0.516 = 1$. The aggregate output at point C is $Q(\bar{q}_{11}) = 4.593 \times 0.516 = 2.37$. The OTE of manager 1 in period 1 is $OTE^1(x_{11}, q_{11}, z_{11}) = Q(q_{11})/Q(\bar{q}_{11}) = 0.422$ (i.e., the aggregate output at point A divided by the aggregate output at point C).

The fact that the OTE of a manager can be defined in terms of aggregate outputs means it can be depicted in input-output space. It can also be viewed as a TFPI. For example, points A and C in Fig. 1.5 map to points A and C in Fig. 1.8. The frontier depicted in this figure is the frontier depicted earlier in Fig. 1.4. The TFP at point A is $TFP(x_{11}, q_{11}) = Q(q_{11})/X(x_{11}) =$ slope 0A $= 1$. The TFP at point

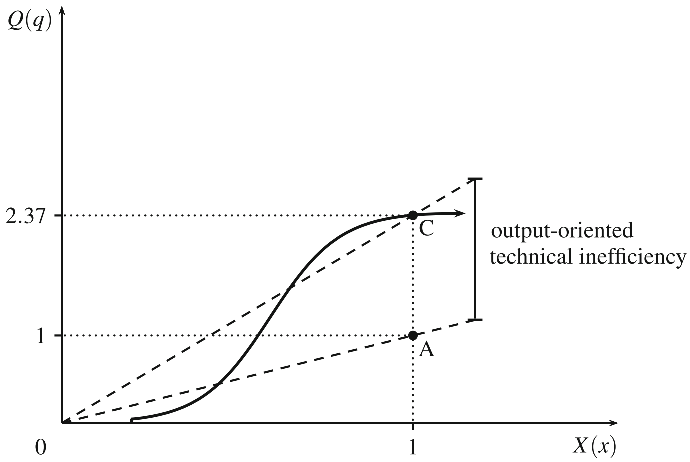

Fig. 1.8 Output-oriented technical inefficiency. The gap between the rays passing through points A and C is due to technical inefficiency

C is $TFP(x_{11}, \bar{q}_{11}) = Q(\bar{q}_{11})/X(x_{11})$ = slope 0C = 2.37. The OTE of manager 1 in period 1 is $OTE^1(x_{11}, q_{11}, z_{11}) = TFP(x_{11}, q_{11})/TFP(x_{11}, \bar{q}_{11}) = 0.422$ (i.e., the TFP at point A divided by the TFP at point C).

1.5.2 Input-Oriented Measures

Input-oriented measures of efficiency are relevant measures of managerial performance in situations where managers have placed nonnegative values on inputs (again, not necessarily market values) and outputs have been predetermined. In these situations, the relevance of a particular measure depends on how easily the manager has been able to choose the input mix. If, for example, the input mix of the firm has been predetermined, then the most relevant measure is input-oriented technical efficiency (ITE). Again, several measures of ITE can be found in the literature. In this book, the ITE of manager i in period t is defined as $ITE^t(x_{it}, q_{it}, z_{it}) = 1/D_I^t(x_{it}, q_{it}, z_{it})$. Equivalently,

$$ITE^t(x_{it}, q_{it}, z_{it}) = X(\bar{x}_{it})/X(x_{it}) \tag{1.14}$$

where $X(x_{it})$ is the aggregate input of the firm and $X(\bar{x}_{it}) = X(x_{it})/D_I^t(x_{it}, q_{it}, z_{it})$ is the minimum aggregate input needed to produce q_{it} in period t when using a scalar multiple of x_{it} in an environment characterised by z_{it}. The right-hand side of (1.14) is an input index. If environmental variables have been predetermined, then it can be viewed as a measure of how well the manager has solved problem (1.11).

For a numerical example, reconsider the input minimisation problem depicted earlier in Fig. 1.6. In that figure, the inputs of firm 1 in period 1 were represented by point A. The aggregate input at point A is $X(x_{11}) = 1.299 \times 0.77 = 1$. The aggregate input at point B is $X(\bar{x}_{11}) = 0.727 \times 0.77 = 0.56$. The ITE of manager 1 in period 1 is $ITE^1(x_{11}, q_{11}, z_{11}) = X(\bar{x}_{11})/X(x_{11}) = 0.56/1 = 0.56$ (i.e., the aggregate input at point B divided by the aggregate input at point A).

The fact that the ITE of a manager can be defined in terms of aggregate inputs means it can also be depicted in input-output space. It can also be viewed as a TFPI. For example, points A and B in Fig. 1.6 map to points A and B in Fig. 1.9. The frontier passing through point B in Fig. 1.9 is the frontier depicted earlier in Figs. 1.4 and 1.8. The TFP at point A is $TFP(x_{11}, q_{11}) = Q(q_{11})/X(x_{11})$ = slope 0A = 1. The TFP at point B is $TFP(\bar{x}_{11}, q_{11}) = Q(q_{11})/X(\bar{x}_{11})$ = slope 0B= 1.786. The ITE of manager 1 in period 1 is $ITE^1(x_{11}, q_{11}, z_{11}) = TFP(x_{11}, q_{11})/TFP(\bar{x}_{11}, q_{11}) = 1/1.786 = 0.56$ (i.e., the TFP at point A divided by the TFP at point B).

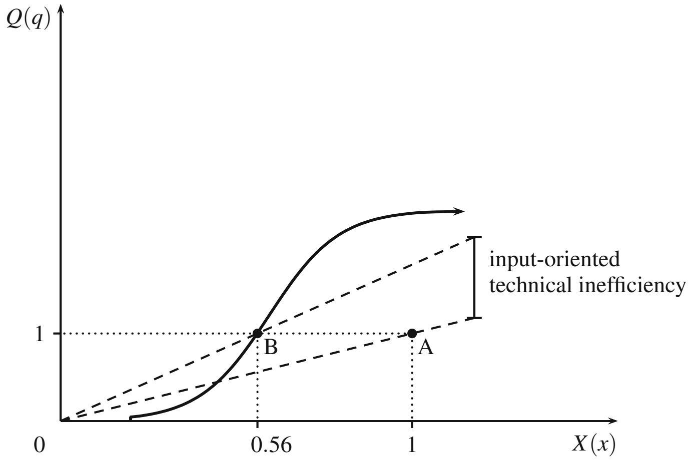

Fig. 1.9 Input-oriented technical inefficiency. The gap between the rays passing through points A and B is due to technical inefficiency

1.5.3 Productivity-Oriented Measures

Productivity-oriented measures of efficiency are relevant measures of managerial performance in situations where managers have placed nonnegative values on outputs and inputs (again, not necessarily market values) and chosen at least one output and at least one input freely. In these situations, the relevance of a particular measure depends on how easily the manager has been able to choose the output mix and the input mix. If, for example, all outputs and inputs have been chosen freely, then the most relevant measure is technical, scale and mix efficiency (TSME). The TSME of manager i in period t is

$$TSME^t(x_{it}, q_{it}, z_{it}) = TFP(x_{it}, q_{it})/TFP^t(z_{it}) \tag{1.15}$$

where $TFP(x_{it}, q_{it}) = Q(q_{it})/X(x_{it})$ is the TFP of the firm and $TFP^t(z_{it})$ is the maximum TFP that is possible in period t in an environment characterised by z_{it}. The right-hand side of (1.15) is a TFPI. If environmental variables have been predetermined, then it can be viewed as a measure of how well the manager has solved problem (1.12).

For a numerical example, reconsider the TFP maximisation problem depicted earlier in Fig. 1.7. Relevant parts of that figure are now reproduced in Fig. 1.10. In these figures, the outputs and inputs of firm 1 in period 1 map to point A. The TFP at point A is $TFP(x_{11}, q_{11}) =$ slope 0A $= 1/1 = 1$. The TFP at any point on the line connecting points N and E is $TFP^1(z_{11}) =$ slope 0E $= 2.744/0.7805 = 3.516$. The TSME of manager 1 in period 1 is $TSME^1(x_{11}, q_{11}, z_{11}) = TFP(x_{11}, q_{11})/TFP^1(z_{11}) = 0.284$

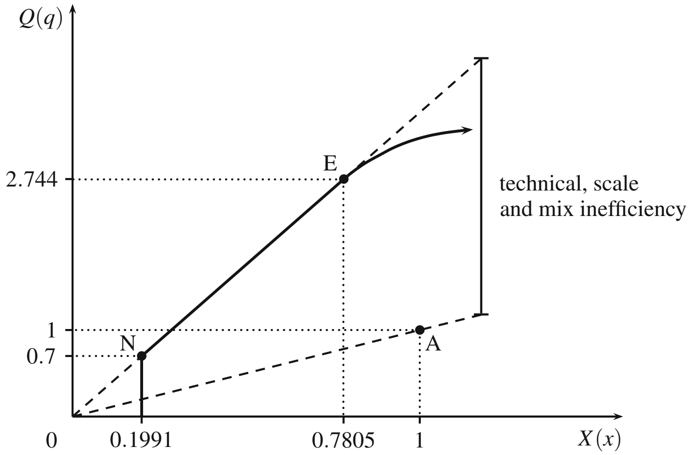

Fig. 1.10 Technical, scale and mix inefficiency. The gap between the rays passing through points A and E is due to technical, scale and mix inefficiency

(i.e., the TFP at point A divided by the TFP at any point on the line connecting points N and E).

The measure of TSME defined by (1.15) can be decomposed into a measure of technical efficiency and a measure of scale and mix efficiency. Both output- and input-oriented decompositions are available. The technical efficiency components are the measures of OTE and ITE defined by (1.13) and (1.14). The scale and mix efficiency components are productivity-oriented measures of economies of scale and substitution. Economies of scale and substitution are the benefits obtained by changing the scale of operations, the output mix, and the input mix. On the output side, the so-called output-oriented scale and mix efficiency (OSME) of manager i in period t is

$$OSME^t(x_{it}, q_{it}, z_{it}) = TFP(x_{it}, \bar{q}_{it})/TFP^t(z_{it}) \tag{1.16}$$

where $TFP(x_{it}, \bar{q}_{it}) = Q(\bar{q}_{it})/X(x_{it})$ is the maximum TFP possible when using x_{it} to produce a scalar multiple of q_{it} in period t in an environment characterised by z_{it}. Equations (1.13), (1.15) and (1.16) imply that

$$OSME^t(x_{it}, q_{it}, z_{it}) = TSME^t(x_{it}, q_{it}, z_{it})/OTE^t(x_{it}, q_{it}, z_{it}). \tag{1.17}$$

This equation says that OSME is the component of TSME that remains after accounting for OTE. On the input side, the so-called input-oriented scale and mix efficiency (ISME) of manager i in period t is

$$ISME^t(x_{it}, q_{it}, z_{it}) = TFP(\bar{x}_{it}, q_{it})/TFP^t(z_{it}) \tag{1.18}$$

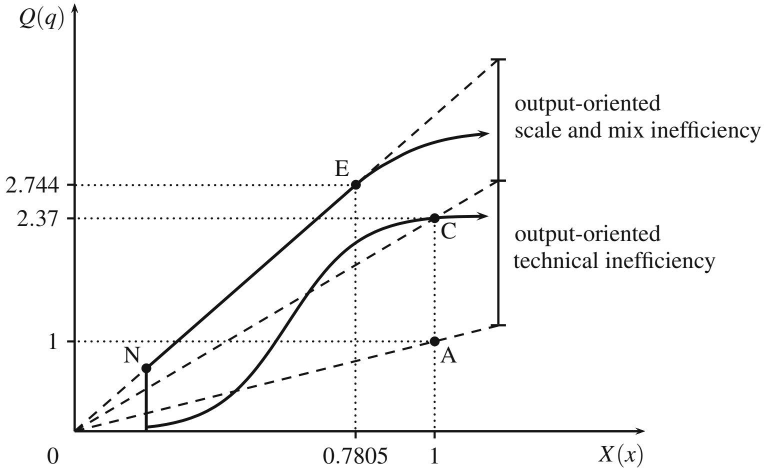

Fig. 1.11 Technical, scale and mix inefficiency. The gap between the rays passing through points A and C is due to technical inefficiency. The gap between the rays passing through points C and E is due to scale and mix inefficiency

where $TFP(\bar{x}_{it}, q_{it}) = Q(q_{it})/X(\bar{x}_{it})$ is the maximum TFP possible when using a scalar multiple of x_{it} to produce q_{it} in period t in an environment characterised by z_{it}. Equations (1.14), (1.15) and (1.18) imply that

$$ISME^t(x_{it}, q_{it}, z_{it}) = TSME^t(x_{it}, q_{it}, z_{it})/ITE^t(x_{it}, q_{it}, z_{it}). \tag{1.19}$$

This equation says that ISME is the component of TSME that remains after accounting for ITE.

For a numerical example, reconsider the measures of OTE, ITE and TSME depicted in Figs. 1.8, 1.9 and 1.10. Relevant parts of those figures are now reproduced in Figs. 1.11 and 1.12. In Fig. 1.11, the OSME of manager 1 in period 1 is $OSME^1(x_{11}, q_{11}, z_{11}) = TFP(x_{11}, \bar{q}_{11})/TFP^1(z_{11}) = 0.674$ (i.e., the TFP at point C divided by the TFP at any point on the line connecting points N and E). In Fig. 1.12, the ISME of manager 1 in period 1 is $ISME^1(x_{11}, q_{11}, z_{11}) = TFP(\bar{x}_{11}, q_{11})/TFP^1(z_{11}) = 0.508$ (i.e., the TFP at point B divided by the TFP at any point on the line connecting points N and E).

1.5.4 Other Measures

Other measures of efficiency discussed in this book include metatechnology ratios and measures of revenue, cost, profit, mix, allocative and scale efficiency. Metatechnology ratios can be viewed as measures of how well managers have chosen their production technologies (i.e., how well they have chosen their 'books of instructions').

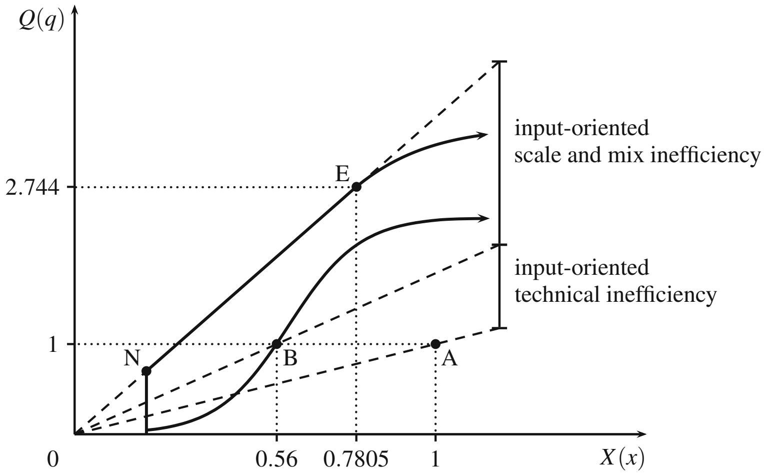

Fig. 1.12 Technical, scale and mix inefficiency. The gap between the rays passing through points A and B is due to technical inefficiency. The gap between the rays passing through points B and E is due to scale and mix inefficiency

Measures of revenue, cost and profit efficiency are measures of how well managers have maximised revenue, minimised cost and maximised profit. Measures of mix efficiency are measures of how well managers have captured economies of substitution (i.e., the benefits obtained by substituting some outputs for others, or by substituting some inputs for others). Measures of scale efficiency are measures of how well managers have captured economies of scale (i.e., the benefits obtained by changing the scale of operations).

1.6 Piecewise Frontier Analysis

Estimating and/or predicting levels of efficiency involves estimating production frontiers. A widely-used estimation approach involves enveloping scatterplots of data points as tightly as possible without violating any assumed properties of production technologies. Some of the most common assumptions lead to estimated frontiers that are comprised of multiple linear segments (or pieces). The associated frontiers are known as piecewise frontiers.[4]

[4]In mathematics, a piecewise function is a function defined on a sequence of intervals (or subdomains). Examples include the absolute value function and the Heaviside step function.

1.6.1 Basic Models

The most common piecewise frontier models (PFMs) are underpinned by the following assumptions:

PF1: production possibilities sets can be represented by distance, revenue, cost and/or profit functions;
PF2: all relevant quantities, prices and environmental variables are observed and measured without error;
PF3: production frontiers are locally (or piecewise) linear;
PF4: outputs, inputs and environmental variables are strongly disposable; and
PF5: production possibilities sets are convex.

If these assumptions are true, then most measures of efficiency can be estimated using linear programming (LP). The associated models and estimators are commonly known as data envelopment analysis (DEA) and estimators.

Output-Oriented Models

Output-oriented PFMs are mainly used to estimate the measure of OTE defined by (1.13). If there are I firms in the dataset and assumptions PF1 to PF5 are true, then the DEA estimation problem can be written as

$$\max_{\mu,\lambda_{11},\ldots,\lambda_{It}} \Big\{ \mu : \ \mu q_{it} \leq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} q_{hr}, \ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} z_{hr} \leq z_{it},$$
$$\sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} x_{hr} \leq x_{it}, \ \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} = 1, \ \lambda_{hr} \geq 0 \text{ for all } h \text{ and } r \Big\}. \quad (1.20)$$

This LP problem seeks to scale up the output vector while holding inputs and environmental variables fixed. The value of μ at the optimum is an estimate of the reciprocal of $OTE^t(x_{it}, q_{it}, z_{it})$.

For a numerical example, reconsider the toy data in Table 1.1. The estimation problem for manager 1 in period 1 is

$$\begin{array}{lllllll}
\max\limits_{\mu,\lambda_{11},\ldots,\lambda_{51}} & \mu & & & & & \\
\text{s.t.} & 1\mu - 1\lambda_{11} & - 1\lambda_{21} & - 2.37\lambda_{31} & - 2.11\lambda_{41} & - 1.81\lambda_{51} & \leq 0 \\
 & 1\mu - 1\lambda_{11} & - 1\lambda_{21} & - 2.37\lambda_{31} & - 2.11\lambda_{41} & - 3.62\lambda_{51} & \leq 0 \\
 & 1\lambda_{11} & + 1\lambda_{21} & + 1\lambda_{31} & + 1\lambda_{41} & + 1\lambda_{51} & \leq 1 \\
 & 1\lambda_{11} & + 0.56\lambda_{21} & + 1\lambda_{31} & + 1.05\lambda_{41} & + 1.05\lambda_{51} & \leq 1 \\
 & 1\lambda_{11} & + 0.56\lambda_{21} & + 1\lambda_{31} & + 0.7\lambda_{41} & + 0.7\lambda_{51} & \leq 1 \\
 & \lambda_{11} & + \lambda_{21} & + \lambda_{31} & + \lambda_{41} & + \lambda_{51} & = 1 \\
\text{and} & \lambda_{11}, \ldots, \lambda_{51} \geq 0. & & & & &
\end{array}$$

The value of the μ at the optimum is 2.37. The associated estimate of OTE is $O\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 1/2.37 = 0.422$. DEA estimates of OTE for other managers

Table 1.7 DEA estimates of OTE[a]

Row	Firm	Period	OTE
A	1	1	0.422
B	2	1	1
C	3	1	1
D	4	1	1
E	5	1	1
F	1	2	0.865
G	2	2	1
H	3	2	1
I	4	2	0.871
J	5	2	1
K	1	3	1
L	2	3	1
M	3	3	0.653
N	4	3	1
O	5	3	0.844
P	1	4	0.594
R	2	4	0.671
S	3	4	0.583
T	4	4	1
U	5	4	0.654
V	1	5	1
W	2	5	0.895
X	3	5	0.836
Y	4	5	0.516
Z	5	5	0.867

[a]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

in other periods can be obtained in a similar way and are reported in Table 1.7. The solution for manager 1 in period 1 is depicted in Fig. 1.13. In this figure, the outputs of firm 1 in period 1 map to point A. The piecewise frontier passing through point C is an estimate of the true frontier depicted earlier in Fig. 1.5.

Input-Oriented Models

Input-oriented PFMs are mainly used to estimate the measure of ITE defined by (1.14). If there are I firms in the dataset and assumptions PF1 to PF5 are true, then the DEA estimation problem can be written as

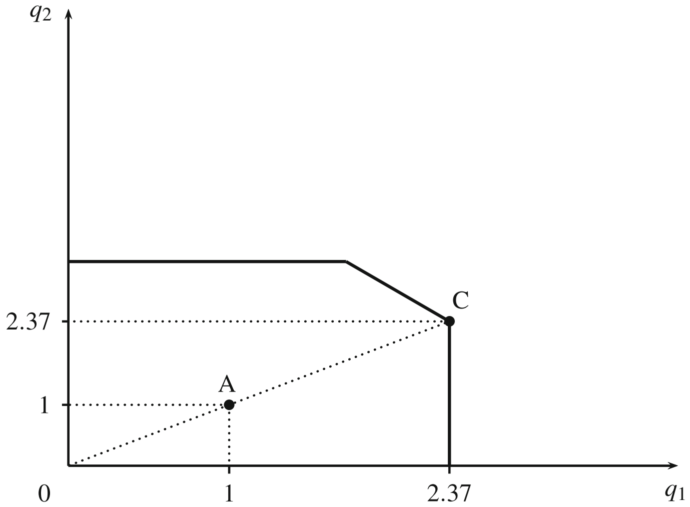

Fig. 1.13 An estimate of output-oriented technical efficiency. In the case of firm A, the DEA estimate of OTE is $O\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 1/2.37 = 0.422$

$$\min_{\mu,\lambda_{11},\ldots,\lambda_{It}} \Big\{\mu : \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} q_{hr} \geq q_{it}, \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} z_{hr} \leq z_{it},$$
$$\mu x_{it} \geq \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} x_{hr}, \sum_{h=1}^{I}\sum_{r=1}^{t} \lambda_{hr} = 1, \ \lambda_{hr} \geq 0 \text{ for all } h \text{ and } r\Big\}. \quad (1.21)$$

This LP problem seeks to scale down the input vector while holding outputs and environmental variables fixed. The value of μ at the optimum is an estimate of $ITE^t(x_{it}, q_{it}, z_{it})$.

For a numerical example, reconsider the toy data in Table 1.1. The estimation problem for firm 1 in period 1 is

$$\begin{array}{llllllll}
\min\limits_{\mu,\lambda_{11},\ldots,\lambda_{It}} & \mu \\
\text{s.t.} & & 1\lambda_{11} & + 1\lambda_{21} & + 2.37\lambda_{31} & + 2.11\lambda_{41} & + 1.81\lambda_{51} & \geq 1 \\
& & 1\lambda_{11} & + 1\lambda_{21} & + 2.37\lambda_{31} & + 2.11\lambda_{41} & + 3.62\lambda_{51} & \geq 1 \\
& & 1\lambda_{11} & + 1\lambda_{21} & + 1\lambda_{31} & + 1\lambda_{41} & + 1\lambda_{51} & \leq 1 \\
& 1\mu & - 1\lambda_{11} & - 0.56\lambda_{21} & - 1\lambda_{31} & - 1.05\lambda_{41} & - 1.05\lambda_{51} & \geq 0 \\
& 1\mu & - 1\lambda_{11} & - 0.56\lambda_{21} & - 1\lambda_{31} & - 0.7\lambda_{41} & - 0.7\lambda_{51} & \geq 0 \\
& & \lambda_{11} & + \lambda_{21} & + \lambda_{31} & + \lambda_{41} & + \lambda_{51} & = 1 \\
\text{and} & \lambda_{11}, \ldots, \lambda_{51} \geq 0.
\end{array}$$

The value of μ at the optimum is $I\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 0.56$. DEA estimates of ITE for other firms in other periods can be obtained in a similar way and are reported in

Table 1.8 DEA estimates of ITE[a]

Row	Firm	Period	ITE
A	1	1	0.56
B	2	1	1
C	3	1	1
D	4	1	1
E	5	1	1
F	1	2	0.954
G	2	2	1
H	3	2	1
I	4	2	0.955
J	5	2	1
K	1	3	1
L	2	3	1
M	3	3	0.604
N	4	3	1
O	5	3	0.777
P	1	4	0.551
R	2	4	0.657
S	3	4	0.669
T	4	4	1
U	5	4	0.689
V	1	5	1
W	2	5	0.846
X	3	5	0.881
Y	4	5	0.387
Z	5	5	0.5

[a]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

Table 1.8. The solution for firm 1 in period 1 is depicted in Fig. 1.14. In this figure, the inputs of firm 1 in period 1 map to point A. The piecewise frontier passing through point B is an estimate of the true frontier depicted earlier in Fig. 1.6.

Productivity-Oriented Models

Productivity-oriented PFMs are mainly used to estimate the measure of TSME defined by (1.15). If estimates of OTE and ITE are available, then Eqs. (1.17) and (1.19) can subsequently be used to estimate the measures of OSME and ISME defined by (1.16) and (1.18).

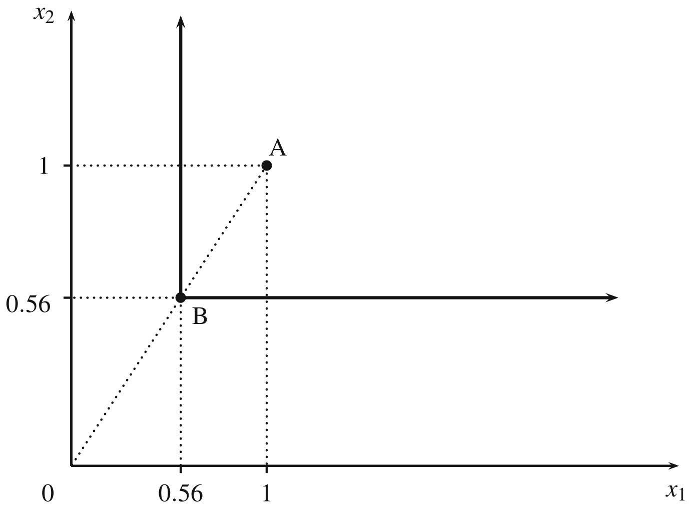

Fig. 1.14 An estimate of input-oriented technical efficiency. In the case of firm A, the DEA estimate of ITE is $I\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 0.56/1 = 0.56$

Estimating the measure of TSME of defined by (1.15) involves estimating $TFP^t(z_{it})$. If there are I firms in the dataset and assumptions PF1 to PF5 are true, then the DEA estimation problem can be written as

$$\max_{q,x,\mu,\theta_{11},\ldots,\theta_{It}} \Bigg\{Q(q) : \; q \leq \sum_{h=1}^{I}\sum_{r=1}^{t}\theta_{hr}q_{hr}, \;\; \sum_{h=1}^{I}\sum_{r=1}^{t}\theta_{hr}z_{hr} \leq \mu z_{it}, \; X(x) = 1,$$
$$\sum_{h=1}^{I}\sum_{r=1}^{t}\theta_{hr}x_{hr} \leq x, \;\; \sum_{h=1}^{I}\sum_{r=1}^{t}\theta_{hr} = \mu, \; \theta_{hr} \geq 0 \text{ for all } h \text{ and } r\Bigg\}. \quad (1.22)$$

If the aggregator functions are linear, then this problem is an LP problem. Whether or not the aggregator functions are linear, the value of the objective function at the optimum is an estimate of $TFP^t(z_{it})$. This can be substituted into (1.15) to obtain an estimate of $TSME^t(x_{it}, q_{it}, z_{it})$.

For a numerical example, reconsider the toy data in Table 1.1. Also suppose that $Q(q) = 0.484q_1 + 0.516q_2$ and $X(x) = 0.23x_1 + 0.77x_2$ (these functions were used earlier to compute the aggregate outputs and inputs in Table 1.2). The estimation problem for firm 1 in period 1 is

$$
\begin{array}{llllllll}
\max\limits_{q,x,\mu,\theta} & 0.484q_1 + 0.516q_2 & & & & & & \\
\text{s.t.} & q_1 - 1\theta_{11} - 1\theta_{21} & - 2.37\theta_{31} & - 2.11\theta_{41} & - 1.81\theta_{51} & & & \leq 0 \\
& q_2 - 1\theta_{11} - 1\theta_{21} & - 2.37\theta_{31} & - 2.11\theta_{41} & - 3.62\theta_{51} & & & \leq 0 \\
& 1\theta_{11} + 1\theta_{21} & + 1\theta_{31} & + 1\theta_{41} & + 1\theta_{51} & - 1\mu & & \leq 0 \\
& & & & 0.23x_1 & + 0.77x_2 & & = 1 \\
& 1\theta_{11} + 0.56\theta_{21} & + 1\theta_{31} & + 1.05\theta_{41} & + 1.05\theta_{51} & - x_1 & & \leq 0 \\
& 1\theta_{11} + 0.56\theta_{21} & + 1\theta_{31} & + 0.7\theta_{41} & + 0.7\theta_{51} & - x_2 & & \leq 0 \\
& \theta_{11} + \theta_{21} & + \theta_{31} & + \theta_{41} & + \theta_{51} & - \mu & & = 0 \\
\text{and} & q_1, q_2, x_1, x_2, \theta_{11}, \ldots, \theta_{51} \geq 0. & & & & & &
\end{array}
$$

The value of the objective function at the optimum is $T\hat{F}P^1(z_{11}) = 3.516$. The TFP of firm 1 in period 1 is $TFP(x_{11}, q_{11}) = 1$. The associated DEA estimate of TSME is $T\hat{S}ME^1(x_{11}, q_{11}, z_{11}) = TFP(x_{11}, q_{11})/T\hat{F}P^1(z_{11}) = 0.284$. DEA estimates of TSME for other managers in other periods can be obtained in a similar way and are reported in Table 1.9. This table also reports estimates of OTE, OSME, ITE and ISME. The OTE and ITE estimates are the ones reported earlier in Tables 1.7 and 1.8. The OSME (resp. ISME) estimates were obtained by dividing the TSME estimates by the OTE (resp. ITE) estimates. The results for manager 1 in period 1 are depicted in Fig. 1.15. In this figure, the piecewise frontier passing through points B and E is an estimate of the true frontier passing through point E in Fig. 1.3. The piecewise frontier passing through points B and C is an estimate of the true frontier passing through points B and C in Fig. 1.4. The outputs and inputs of firm 1 in period 1 map to point A. The TFP-maximising point is point E. Points B and C are technically efficient points. The dashed lines passing through these points are iso-productivity rays with different slopes. The DEA estimates of TSME, OSME and ISME for manager 1 in period 1 are given by the ratios of these slopes.

Other Models

Other PFMs discussed in this book include revenue-, cost- and profit-oriented models. These models are mainly used to estimate measures of revenue, cost, profit, allocative, pure mix and pure scale efficiency.

1.6.2 Productivity Analysis

Productivity analysis involves both measuring and explaining changes in productivity. This section focuses on explaining changes in TFP. This involves decomposing proper TFPI numbers into measures of environmental change, technical change, and efficiency change. If production frontiers are piecewise linear, then the easiest way to proceed is to rewrite (1.15) as $TFP(x_{it}, q_{it}) = TFP^t(z_{it}) \times TSME^t(x_{it}, q_{it}, z_{it})$. A

Table 1.9 DEA estimates of TSME, OSME, OTE, ISME and ITE[a,b]

Row	Firm	Period	TSME	OTE	OSME	ITE	ISME
A	1	1	0.284	0.422	0.674	0.56	0.508
B	2	1	0.508	1	0.508	1	0.508
C	3	1	0.674	1	0.674	1	0.674
D	4	1	0.769	1	0.769	1	0.769
E	5	1	1	1	1	1	1
F	1	2	0.602	0.865	0.696	0.954	0.631
G	2	2	1	1	1	1	1
H	3	2	0.999	1	0.999	1	0.999
I	4	2	0.761	0.871	0.874	0.955	0.797
J	5	2	0.899	1	0.899	1	0.899
K	1	3	0.885	1	0.885	1	0.885
L	2	3	0.785	1	0.785	1	0.785
M	3	3	0.534	0.653	0.819	0.604	0.886
N	4	3	1	1	1	1	1
O	5	3	0.569	0.844	0.674	0.777	0.732
P	1	4	0.547	0.594	0.921	0.551	0.992
R	2	4	0.578	0.671	0.861	0.657	0.880
S	3	4	0.429	0.583	0.737	0.669	0.642
T	4	4	0.811	1	0.811	1	0.811
U	5	4	0.607	0.654	0.928	0.689	0.881
V	1	5	0.896	1	0.896	1	0.896
W	2	5	0.619	0.895	0.692	0.846	0.732
X	3	5	0.566	0.836	0.677	0.881	0.643
Y	4	5	0.384	0.516	0.745	0.387	0.994
Z	5	5	0.5	0.867	0.577	0.5	1

[a]TSME = OTE × OSME = ITE × ISME. Some estimates may be incoherent at the third decimal place due to rounding (e.g., the product of the OTE and OSME estimates in row Z is not exactly equal to 0.5 due to rounding)
[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

similar equation holds for firm k in period s. Substituting these equations into (1.9) yields

$$TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = TFP^t(z_{it})/TFP^s(z_{ks}) \times TSME^t(x_{it}, q_{it}, z_{it})/TSME^s(x_{ks}, q_{ks}, z_{ks}). \quad (1.23)$$

The first ratio on the right-hand side is an environment and technology index (ETI) (i.e., a combined measure of environmental and technical change). The second ratio is a technical, scale and mix efficiency index (TSMEI).

Output- and input-oriented decompositions of TFPI numbers are also available. For an output-oriented decomposition, the easiest way to proceed is to rewrite

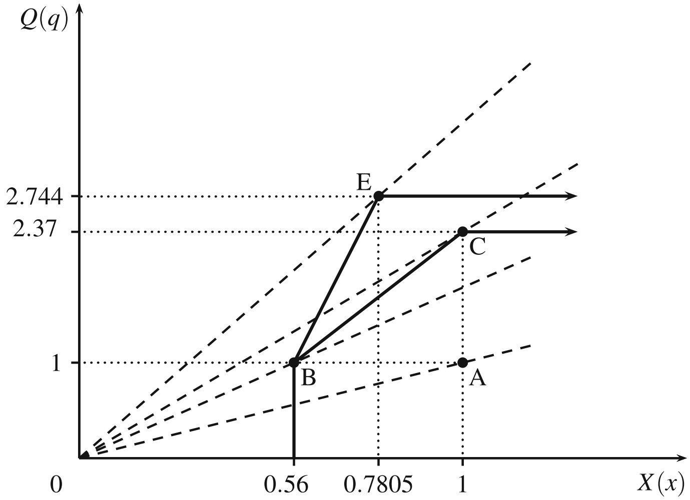

Fig. 1.15 Estimates of technical, scale and mix efficiency. The DEA estimates of TSME, OSME and ISME for manager 1 in period 1 are $T\hat{S}ME^1(x_{11}, q_{11}, z_{11})$ = (slope 0A)/(slope 0E) = 0.2844, $O\hat{S}ME^1(x_{11}, q_{11}, z_{11})$ = (slope 0C)/(slope 0E) = 0.674 and $I\hat{S}ME^1(x_{11}, q_{11}, z_{11})$ = (slope 0B)/(slope 0E) = 0.508

(1.17) as $TSME^t(x_{it}, q_{it}, z_{it}) = OTE^t(x_{it}, q_{it}, z_{it}) \times OSME^t(x_{it}, q_{it}, z_{it})$. A similar equation holds for firm k in period s. Substituting these equations into (1.23) yields

$$\begin{aligned} TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = {} & TFP^t(z_{it})/TFP^s(z_{ks}) \\ & \times OTE^t(x_{it}, q_{it}, z_{it})/OTE^s(x_{ks}, q_{ks}, z_{ks}) \\ & \times OSME^t(x_{it}, q_{it}, z_{it})/OSME^s(x_{ks}, q_{ks}, z_{ks}). \end{aligned} \quad (1.24)$$

The first ratio on the right-hand side is the ETI in (1.23). The second ratio is an output-oriented technical efficiency index (OTEI). The last ratio is an output-oriented scale and mix efficiency index (OSMEI).

For an input-oriented decomposition, the easiest way to proceed is to rewrite (1.19) as $TSME^t(x_{it}, q_{it}, z_{it}) = ITE^t(x_{it}, q_{it}, z_{it}) \times ISME^t(x_{it}, q_{it}, z_{it})$. A similar equation holds for firm k in period s. Substituting these equations into (1.23) yields

$$\begin{aligned} TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = {} & TFP^t(z_{it})/TFP^s(z_{ks}) \\ & \times ITE^t(x_{it}, q_{it}, z_{it})/ITE^s(x_{ks}, q_{ks}, z_{ks}) \\ & \times ISME^t(x_{it}, q_{it}, z_{it})/ISME^s(x_{ks}, q_{ks}, z_{ks}). \end{aligned} \quad (1.25)$$

The first ratio on the right-hand side is the ETI in (1.23) and (1.24). The second ratio is an input-oriented technical efficiency index (ITEI). The last ratio is an input-oriented scale and mix efficiency index (ISMEI).

For a numerical example, reconsider the toy data in Tables 1.1 and 1.2. Associated Lowe TFPI numbers were reported earlier in column L of Table 1.6. Associated DEA estimates of OTE, OSME, ITE and ISME were reported earlier in Table 1.9. Output- and input-oriented decompositions of the TFPI numbers are now reported in Table 1.10. The OTEI, OSMEI, ITEI and ISMEI numbers were obtained by dividing the estimates of OTE, OSME, ITE and ISME for each firm in each period by the corresponding estimates for firm 1 in period 1. The ETI numbers were obtained as residuals (i.e., ETI = TFPI/(OTEI×OSMEI) = TFPI/(ITEI×ISMEI)).

Table 1.10 Output- and input-oriented decompositions of Lowe TFPI numbers using DEA[a,b]

Firm	Period	TFPI	ETI	OTEI	OSMEI	ETI	ITEI	ISMEI
1	1	1	1	1	1	1	1	1
2	1	1.786	1	2.37	0.753	1	1.786	1
3	1	2.37	1	2.37	1	1	1.786	1.327
4	1	2.703	1	2.37	1.141	1	1.786	1.514
5	1	3.516	1	2.37	1.483	1	1.786	1.969
1	2	2.117	1	2.05	1.033	1	1.704	1.243
2	2	3.515	1	2.37	1.483	1	1.786	1.968
3	2	3.513	1	2.37	1.482	1	1.786	1.967
4	2	2.675	1	2.063	1.297	1	1.705	1.569
5	2	3.159	1	2.37	1.333	1	1.786	1.769
1	3	3.110	1.000	2.370	1.312	1.000	1.786	1.742
2	3	2.760	1.000	2.370	1.165	1.000	1.786	1.546
3	3	1.879	1.000	1.547	1.215	1.000	1.078	1.743
4	3	3.516	1.000	2.370	1.483	1.000	1.786	1.969
5	3	2	1.000	2.000	1.000	1.000	1.388	1.441
1	4	1.923	1.000	1.408	1.366	1.000	0.984	1.954
2	4	2.032	1.000	1.590	1.278	1.000	1.173	1.732
3	4	1.509	1.000	1.381	1.093	1.000	1.195	1.263
4	4	2.852	1.000	2.370	1.203	1.000	1.786	1.597
5	4	2.134	1.000	1.550	1.376	1.000	1.230	1.735
1	5	3.150	1.000	2.370	1.329	1.000	1.786	1.764
2	5	2.176	1.000	2.120	1.026	1.000	1.510	1.441
3	5	1.991	1.000	1.982	1.004	1.000	1.573	1.266
4	5	1.351	1.000	1.223	1.105	1.000	0.691	1.956
5	5	1.758	1.000	2.054	0.856	1.000	0.893	1.969

[a]TFPI = ETI × OTEI × OSMEI = ETI × ITEI × ISMEI. Some index numbers may be incoherent at the third decimal place due to rounding (e.g., the product of the ETI, OTEI and OSMEI numbers in row 2 is not exactly equal to the TFPI number due to rounding)
[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

1.6.3 Other Models

Other PFMs discussed in this book include free disposal hull (FDH) and metafrontier models. FDH models are obtained by relaxing the assumption that production possibilities sets are convex. Metafrontier models can be used to decompose measures of technical efficiency into metatechnology ratios and associated measures of residual technical efficiency.

1.7 Deterministic Frontier Analysis

Production frontiers are often represented by distance, revenue, cost and/or profit functions. These functions can sometimes be written in the form of regression models in which the explanatory variables are deterministic (i.e., not random). The associated frontiers are known as deterministic frontiers.

1.7.1 Basic Models

Deterministic frontier models (DFMs) are underpinned by the following assumptions:

DF1 production possibilities sets can be represented by distance, revenue, cost and/or profit functions;

DF2 all relevant quantities, prices and environmental variables are observed and measured without error; and

DF3 the functional forms of relevant functions are known.

If these assumptions are true, then production frontiers can be estimated using single-equation regression models with error terms representing inefficiency.

Output-Oriented Models

Output-oriented DFMs are mainly used to estimate the measure of OTE defined by (1.13). This involves estimating the output distance function. Output distance functions can be written in the form of regression models with nonnegative errors representing output-oriented technical inefficiency. For example, consider the following output distance function:

$$D_O^t(x_{it}, q_{it}, z_{it}) = \left(A(t) \prod_{j=1}^{J} z_{jit}^{\delta_j} \prod_{m=1}^{M} x_{mit}^{\beta_m} \right)^{-1} \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau} \right)^{1/\tau} \tag{1.26}$$

where $A(t) > 0, A(t) \geq A(t-1), \beta = (\beta_1, \ldots, \beta_M)' \geq 0, \gamma = (\gamma_1, \ldots, \gamma_N)' \geq 0, \tau \geq 1$ and $\gamma'\iota = 1$. After some simple algebra, this function can be rewritten as

$$\ln q_{1it} = \alpha(t) + \sum_{j=1}^{J} \delta_j \ln z_{jit} + \sum_{m=1}^{M} \beta_m \ln x_{mit} - \frac{1}{\tau} \ln \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{*\tau} \right) - u_{it} \quad (1.27)$$

where $\alpha(t) \equiv \ln A(t)$ is an output-oriented measure of technical change, $q_{nit}^* \equiv q_{nit}/q_{1it}$ denotes a normalised output, and $u_{it} \equiv -\ln OTE^t(x_{it}, q_{it}, z_{it}) \geq 0$ denotes an output-oriented technical inefficiency effect.

Input-Oriented Models

Input-oriented DFMs are mainly used to estimate the measure of ITE defined by (1.14). This involves estimating the input distance function. Input distance functions can be written in the form of regression models with nonnegative errors representing *input*-oriented technical inefficiency. For example, if the output distance function is given by (1.26), then the input distance function is

$$D_I^t(x_{it}, q_{it}, z_{it}) = \left(B(t) \prod_{j=1}^{J} z_{jit}^{\kappa_j} \prod_{m=1}^{M} x_{mit}^{\lambda_m} \right) \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau} \right)^{-1/(\tau\eta)} \quad (1.28)$$

where $\eta = \beta'\iota$, $B(t) = A(t)^{1/\eta}$, $\kappa_j = \delta_j/\eta$ and $\lambda_m = \beta_m/\eta$. After some simple algebra, this function can be rewritten as

$$-\ln x_{1it} = \xi(t) + \sum_{j=1}^{J} \kappa_j \ln z_{jit} + \sum_{m=2}^{M} \lambda_m \ln x_{mit}^* - \frac{1}{\tau\eta} \ln \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau} \right) - u_{it} \quad (1.29)$$

where $\xi(t) \equiv \ln B(t)$ is an input-oriented measure of technical change, $x_{mit}^* \equiv x_{mit}/x_{1it}$ denotes a normalised input, and, in a slight abuse of notation, $u_{it} \equiv -\ln ITE^t(x_{it}, q_{it}, z_{it}) \geq 0$ now denotes an input-oriented technical inefficiency effect.

Other Models

DFMs can also be used to estimate measures of revenue, cost and profit efficiency. This involves estimating revenue, cost and profit functions. These functions can also be written in the form of regression models with nonnegative errors representing inefficiency.

1.7.2 Least Squares Estimation

Least squares (LS) estimation of DFMs involves choosing the unknown parameters to minimise the sum of squared inefficiency effects. In the efficiency literature, it is common to assume that u_{it} is a random variable with the following properties:

LS1 $E(u_{it}) = \mu \geq 0$ for all i and t.
LS2 $var(u_{it}) \propto \sigma_u^2$ for all i and t.
LS3 $cov(u_{it}, u_{ks}) = 0$ if $i \neq k$ or $t \neq s$.
LS4 u_{it} is uncorrelated with the explanatory variables.

LS1 says the inefficiency effects have the same mean. LS2 says they are homoskedastic. LS3 says they are serially and spatially uncorrelated. LS4 is self-explanatory.

If a DFM contains an intercept and LS1 to LS4 are true, then LS estimators for the slope parameters are consistent. A consistent estimator for the intercept can be obtained by adjusting the LS estimator for the intercept upwards by an amount equal to the maximum of the LS residuals. In this book, the associated estimators for the intercept and slope parameters are collectively referred to as corrected least squares (CLS) estimators. In practice, it is common to impose restrictions on the parameters so that the estimated frontier is consistent with any assumed properties of production technologies. If the restrictions are true, then associated restricted least squares (RLS) estimators for the slope parameters are consistent. Again, a consistent estimator for the intercept can be obtained by adjusting the RLS estimator for the intercept upwards by an amount equal to the largest RLS residual. In this book, the associated estimators for the intercept and slope parameters are collectively referred to as corrected restricted least squares (CRLS) estimators.

For a numerical example, reconsider the toy data in Table 1.1. These data have been used to obtain CLS and CRLS estimates of the parameters in (1.27). The estimates are reported in Table 1.11. The CRLS estimates were obtained by restricting $\alpha(t) \geq \alpha(t-1)$, $\beta = (\beta_1, \ldots, \beta_M)' \geq 0$ and $\tau \geq 1$. The CRLS estimates have been used to predict levels of OTE and ITE. The predictions are reported in Table 1.12. The OTE

Table 1.11 LS parameter estimates

Parameter	CLS	CRLS
$\alpha(1) \equiv \ln A(1)$	0.954	1.159
$\alpha(2) \equiv \ln A(2)$	0.903	1.159
$\alpha(3) \equiv \ln A(3)$	0.702	1.159
$\alpha(4) \equiv \ln A(4)$	0.723	1.159
$\alpha(5) \equiv \ln A(5)$	0.782	1.159
δ_1	0.188	−0.056
β_1	0.093	0.280
β_2	0.259	0
γ_1	0.771	0.724
γ_2	0.229	0.276
τ	−0.083	1

Table 1.12 CRLS predictions of OTE and ITE[a]

Row	Firm	Period	OTE	ITE
A	1	1	0.314	0.016
B	2	1	0.369	0.029
C	3	1	0.744	0.348
D	4	1	0.653	0.219
E	5	1	0.715	0.302
F	1	2	0.327	0.018
G	2	2	0.660	0.226
H	3	2	0.938	0.795
I	4	2	0.900	0.686
J	5	2	1	1
K	1	3	0.724	0.315
L	2	3	0.546	0.115
M	3	3	0.447	0.057
N	4	3	0.300	0.014
O	5	3	0.652	0.218
P	1	4	0.353	0.024
R	2	4	0.487	0.077
S	3	4	0.271	0.009
T	4	4	0.790	0.430
U	5	4	0.314	0.016
V	1	5	0.675	0.245
W	2	5	0.668	0.237
X	3	5	0.293	0.013
Y	4	5	0.341	0.022
Z	5	5	0.589	0.151

[a]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

predictions were obtained by evaluating (1.26). The ITE predictions were obtained by evaluating the reciprocal of (1.28). The predictions for manager 1 in period 1 are depicted in Figs. 1.16 and 1.17. In Fig. 1.16 (resp. 1.17), the outputs (resp. inputs) of firm 1 in period 1 map to point A. In Fig. 1.16, the frontier passing through point A* is an estimate of the true frontier depicted earlier in Fig. 1.5. In Fig. 1.17, the frontier passing through point B* is an estimate of the true frontier depicted earlier in Fig. 1.6.

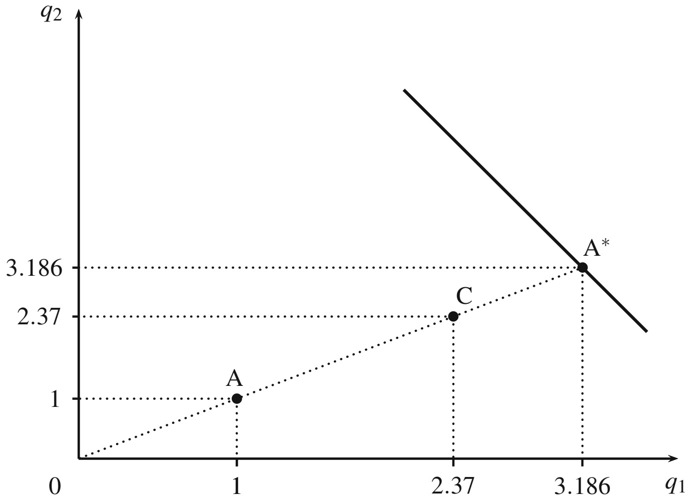

Fig. 1.16 A prediction of output-oriented technical efficiency. In the case of firm A, the CRLS prediction of OTE is $O\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 1/3.186 = 0.314$

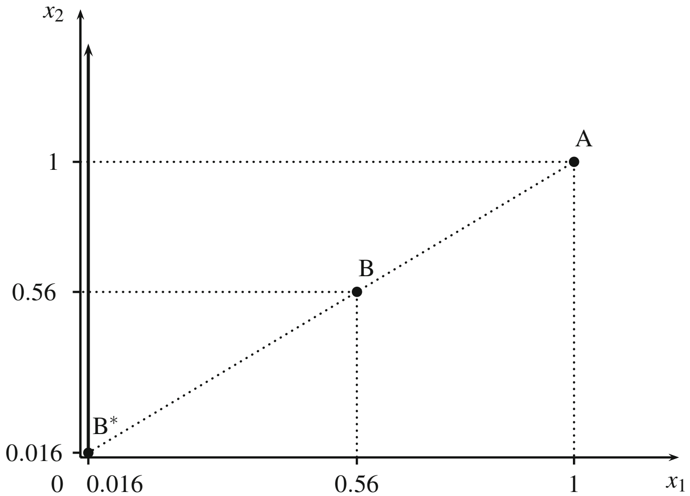

Fig. 1.17 A prediction of input-oriented technical efficiency. In the case of firm A, the CRLS prediction of ITE is $I\hat{T}E^1(x_{11}, q_{11}, z_{11}) = 0.016/1 = 0.016$

1.7.3 Productivity Analysis

For purposes of comparison with Sect. 1.6.2, this section focuses on decomposing proper TFPI numbers. Again, both output- and input-oriented decompositions are available.

For an output-oriented decomposition, a relatively easy way to proceed is to write $TFP(x_{it}, q_{it}) = TFP(x_{it}, q_{it}) \exp(-u_{it})/D_O^t(x_{it}, q_{it}, z_{it})$ where u_{it} denotes an output-oriented technical inefficiency effect. The precise form of this equation depends partly on the form of the output distance function. If the output distance function is given by (1.26), for example, then

$$TFP(x_{it}, q_{it}) = A(t)\left[\prod_{j=1}^{J} z_{jit}^{\delta_j}\right] \times \left[TFP(x_{it}, q_{it}) \prod_{m=1}^{M} x_{mit}^{\beta_m} \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau}\right)^{-1/\tau}\right] \exp(-u_{it}).$$

A similar equation holds for firm k in period s. Substituting these equations into (1.9) yields

$$TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = \left[\frac{A(t)}{A(s)}\right]\left[\prod_{j=1}^{J}\left(\frac{z_{jit}}{z_{jks}}\right)^{\delta_j}\right] \times \left[TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) \prod_{m=1}^{M}\left(\frac{x_{mit}}{x_{mks}}\right)^{\beta_m}\left(\frac{\sum_n \gamma_n q_{nks}^{\tau}}{\sum_n \gamma_n q_{nit}^{\tau}}\right)^{1/\tau}\right] \times \left[\frac{\exp(-u_{it})}{\exp(-u_{ks})}\right]. \tag{1.30}$$

The first term on the right-hand side is an output-oriented technology index (OTI) (i.e., a measure of technical change). The second term is an output-oriented environment index (OEI) (i.e., a measure of environmental change). The third term is an output-oriented scale and mix efficiency index (OSMEI). The last term is an output-oriented technical efficiency index (OTEI). If there are no environmental variables involved in the production process, then the second term vanishes. The conditions under which other terms vanish is left as an exercise for the reader.

For an input-oriented decomposition, a relatively easy way to proceed is to write $TFP(x_{it}, q_{it}) = TFP(x_{it}, q_{it}) D_I^t(x_{it}, q_{it}, z_{it}) \exp(-u_{it})$ where u_{it} now denotes an *input*-oriented technical inefficiency effect. Again, the precise form of this equation depends partly on the form of the distance function. If the input distance function is given by (1.28), for example, then

$$TFP(x_{it}, q_{it}) = B(t)\left[\prod_{j=1}^{J} z_{jit}^{\kappa_j}\right] \times \left[TFP(x_{it}, q_{it}) \prod_{m=1}^{M} x_{mit}^{\lambda_m} \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau}\right)^{-1/(\tau\eta)}\right] \exp(-u_{it}).$$

A similar equation holds for firm k in period s. Substituting these equations into (1.9) yields

$$TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = \left[\frac{B(t)}{B(s)}\right]\left[\prod_{j=1}^{J}\left(\frac{z_{jit}}{z_{jks}}\right)^{\kappa_j}\right] \times \left[TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) \prod_{m=1}^{M}\left(\frac{x_{mit}}{x_{mks}}\right)^{\lambda_m}\left(\frac{\sum_n \gamma_n q_{nks}^{\tau}}{\sum_n \gamma_n q_{nit}^{\tau}}\right)^{1/(\tau\eta)}\right] \times \left[\frac{\exp(-u_{it})}{\exp(-u_{ks})}\right]. \quad (1.31)$$

The first term on the right-hand side is an input-oriented technology index (ITI). The second term is an input-oriented environment index (IEI). The third term is an input-oriented scale and mix efficiency index (ISMEI). The last term is an input-oriented technical efficiency index (ITEI). Again, the conditions under which these terms vanish is left as an exercise for the reader.

For a numerical example, reconsider the toy data in Tables 1.1 and 1.2. Associated Lowe TFPI numbers were reported earlier in Table 1.6. Output- and input-oriented decompositions of these numbers are now reported in Table 1.13. The OTI, OEI, OSMEI, ITI, IEI and ISMEI numbers in each row were obtained by using the CRLS estimates reported in Table 1.11 to evaluate the relevant terms in (1.30) and (1.31). The OTEI and ITEI numbers were obtained as residuals (i.e., OTEI = TFPI/(OTI×OEI×OSMEI) and ITEI = TFPI/(ITI×IEI×ISMEI); these numbers could also have been obtained by taking ratios of the CRLS estimates of OTE and ITE reported earlier in Table 1.12).

1.7.4 Other Models

Other DFMs discussed in this book include various systems of equations. These systems can be used to explain variations in metafrontiers, output supplies and input demands.

Table 1.13 Output- and input-oriented decompositions of Lowe TFPI numbers using CRLS[a,b]

Firm	Period	TFPI	OTI	OEI	OTEI	OSMEI	ITI	IEI	ITEI	ISMEI
1	1	1	1	1	1	1	1	1	1	1
2	1	1.786	1	1	1.176	1.518	1	1	1.786	1
3	1	2.37	1	1	2.37	1	1	1	21.773	0.109
4	1	2.703	1	1	2.081	1.299	1	1	13.695	0.197
5	1	3.516	1	1	2.278	1.543	1	1	18.905	0.186
1	2	2.117	1	0.962	1.041	2.114	1	0.871	1.153	2.108
2	2	3.515	1	0.962	2.102	1.738	1	0.871	14.181	0.285
3	2	3.513	1	1	2.988	1.176	1	1	49.810	0.071
4	2	2.675	1	0.962	2.867	0.970	1	0.871	42.949	0.072
5	2	3.159	1	1	3.186	0.991	1	1	62.651	0.050
1	3	3.110	1	1	2.306	1.349	1	1	19.734	0.158
2	3	2.760	1	1	1.739	1.587	1	1	7.213	0.383
3	3	1.879	1	1	1.426	1.318	1	1	3.546	0.530
4	3	3.516	1	1	0.955	3.682	1	1	0.848	4.145
5	3	2	1	0.962	2.079	1	1	0.871	13.639	0.168
1	4	1.923	1	1	1.125	1.710	1	1	1.522	1.264
2	4	2.032	1	1	1.552	1.309	1	1	4.801	0.423
3	4	1.509	1	0.962	0.864	1.815	1	0.871	0.594	2.918
4	4	2.852	1	0.962	2.516	1.178	1	0.871	26.969	0.121
5	4	2.134	1	1	1	2.134	1	1	1	2.134
1	5	3.150	1	1	2.149	1.465	1	1	15.363	0.205
2	5	2.176	1	0.962	2.129	1.063	1	0.871	14.841	0.168
3	5	1.991	1	0.962	0.934	2.215	1	0.871	0.784	2.915
4	5	1.351	1	1	1.088	1.242	1	1	1.351	1
5	5	1.758	1	1	1.876	0.937	1	1	9.453	0.186

[a]TFPI = OTI × OEI × OTEI × OSMEI = ITI × IEI × ITEI × ISMEI. Some index numbers may be incoherent at the third decimal place due to rounding (e.g., the product of the OTI, OEI, OTEI and OSMEI numbers in row 2 is not exactly equal to the TFPI number due to rounding)
[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

1.8 Stochastic Frontier Analysis

Distance, revenue, cost and profit functions can always be written in the form of regression models with unobserved error terms representing statistical noise and different types of inefficiency. In practice, the noise components are almost always assumed to be stochastic. The associated frontiers are known as stochastic frontiers.

1.8.1 Basic Models

Stochastic frontier models (SFMs) are underpinned by only one assumption, namely that production possibilities sets can be represented by distance, revenue, cost and/or profit functions.

Output-Oriented Models

Output-oriented SFMs are mainly used to estimate the measure of OTE defined by (1.13). This involves estimating the output distance function. Any output distance function can be written in the form of a regression model with an error representing statistical noise and another error representing output-oriented technical inefficiency. For example, any output distance function can be written as

$$\ln q_{1it} = \alpha + \lambda t + \sum_{j=1}^{J} \delta_j \ln z_{jit} + \sum_{m=1}^{M} \beta_m \ln x_{mit} - \sum_{n=1}^{N} \gamma_n \ln q^*_{nit} + v_{it} - u_{it} \tag{1.32}$$

where $q^*_{nit} \equiv q_{nit}/q_{1it}$ is a normalised output, $\gamma = (\gamma_1, \ldots, \gamma_N)'$ is a vector of parameters that sum to one, v_{it} is an error representing statistical noise, and $u_{it} \equiv -\ln D^t_O(x_{it}, q_{it}, z_{it})$ is a nonnegative output-oriented technical inefficiency effect. The exact nature of the noise component depends on the unknown output distance function. If the output distance function is given by (1.26), for example, then

$$v_{it} = [\alpha(t) - \alpha - \lambda t] + \left[\sum_{n=1}^{N} \gamma_n \ln q^*_{nit} - \frac{1}{\tau} \ln \left(\sum_{n=1}^{N} \gamma_n q^{*\tau}_{nit} \right) \right]. \tag{1.33}$$

These terms can be viewed as functional form errors.

Input-Oriented Models

Input-oriented SFMs are mainly used to estimate the measure of ITE defined by (1.14). This involves estimating the input distance function. Any input distance function can be written in the form of a regression model with an error representing statistical noise and another error representing *input*-oriented technical inefficiency. For example, any input distance function can be written as

$$-\ln x_{1it} = \xi(t) + \sum_{m=1}^{M} \lambda_m \ln x^*_{mit} - \sum_{n=1}^{N} \phi_n \ln q_{nit} + v_{it} - u_{it} \tag{1.34}$$

where $x^*_{mit} \equiv x_{mit}/x_{1it}$ is a normalised input, $\lambda = (\lambda_1, \ldots, \lambda_M)'$ is a vector of parameters that sum to one, v_{it} is an error representing statistical noise, and $u_{it} \equiv -\ln ITE^t(x_{it}, q_{it}, z_{it})$ is now a nonnegative input-oriented technical inefficiency effect. In this case, the exact nature of the noise component depends on the unknown input distance function. If the input distance function is given by (1.28), for example, then

$$v_{it} = \left[\sum_{j=1}^{J} \kappa_j \ln z_{jit} \right] + \left[\sum_{n=1}^{N} \phi_n \ln q_{nit} - \frac{1}{\tau\eta} \ln \left(\sum_{n=1}^{N} \gamma_n q_{nit}^{\tau} \right) \right]. \tag{1.35}$$

The first term is an omitted variable error. The second term can be viewed as a functional form error.

Other Models

Revenue-, cost- and profit-oriented SFMs are also available. These models are mainly used to estimate measures of revenue, cost and profit efficiency. This involves estimating revenue, cost and profit functions. These functions can also be written in the form of regression models with error terms representing statistical noise and different types of inefficiency.

1.8.2 Maximum Likelihood Estimation

Maximum likelihood (ML) estimation of SFMs involves choosing the unknown parameters to maximise the joint density (or 'likelihood') of the observed data. For simplicity, consider the output-oriented model defined by (1.32). This model can be written more compactly as

$$y_{it} = \alpha + \lambda t + \sum_{j=1}^{J} \delta_j \ln z_{jit} + \sum_{m=1}^{M} \beta_m \ln x_{mit} - \sum_{n=1}^{N} \gamma_n \ln q_{nit}^{*} + \epsilon_{it} \tag{1.36}$$

where $y_{it} \equiv \ln q_{1it}$ denotes the logarithm of the first output and $\epsilon_{it} \equiv v_{it} - u_{it}$ is a composite error representing statistical noise and output-oriented technical inefficiency. The likelihood of the observed data depends on the assumed probability distributions of v_{it} and u_{it}. It is common to assume that

ML3 v_{it} is an independent $N(0, \sigma_v^2)$ random variable, and
ML4 u_{it} is an independent $N^+(\mu, \sigma_u^2)$ random variable.

If these assumptions are true, then the ML estimators for the unknown parameters in the model are consistent, asymptotically efficient, and asymptotically normal. Following estimation, ML predictions of u_{it} can be obtained by using the ML parameter estimates to evaluate

$$E(u_{it}|\epsilon_{it}) = \mu_{it}^{*} + \sigma_{*} \left(\frac{\phi(\mu_{it}^{*}/\sigma_{*})}{\Phi(\mu_{it}^{*}/\sigma_{*})} \right) \tag{1.37}$$

where $\mu_{it}^{*} \equiv (\mu\sigma_v^2 - \epsilon_{it}\sigma_u^2)/(\sigma_v^2 + \sigma_u^2)$ and $\sigma_{*}^2 \equiv \sigma_v^2\sigma_u^2/(\sigma_v^2 + \sigma_u^2)$. Let $\tilde{u}_{it}$ denote the ML predictor for u_{it}. The associated predictor for OTE is $\exp(-\tilde{u}_{it})$.

Table 1.14 ML parameter estimates

Parameter	ML	RML
α	0.990	0.480
λ	−0.069	0
δ_1	0.258	0.110
β_1	0.148	0.092
β_2	0.279	0.289
γ_1	0.682	0.676
γ_2	0.318	0.324
σ_u^2	0.225	0.000
σ_v^2	0.000	0.083
μ	−0.139	−0.026

Table 1.15 ML predictions of OTE

Row	Firm	Period	ML	RML
A	1	1	0.398	0.995
B	2	1	0.510	0.995
C	3	1	0.944	0.995
D	4	1	0.921	0.995
E	5	1	0.985	0.995
F	1	2	0.492	0.995
G	2	2	0.880	0.995
H	3	2	1.000	0.995
I	4	2	0.381	0.995
J	5	2	0.956	0.995
K	1	3	0.933	0.995
L	2	3	0.511	0.995
M	3	3	0.620	0.995
N	4	3	0.628	0.995
O	5	3	0.765	0.995
P	1	4	0.640	0.995
R	2	4	0.694	0.995
S	3	4	0.528	0.995
T	4	4	0.972	0.995
U	5	4	0.679	0.995
V	1	5	0.884	0.995
W	2	5	0.910	0.995
X	3	5	0.637	0.995
Y	4	5	0.597	0.995
Z	5	5	0.965	0.995

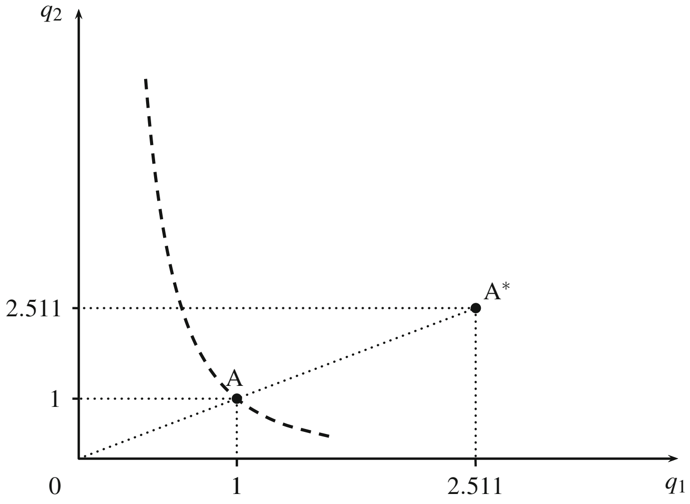

Fig. 1.18 A prediction of output-oriented technical efficiency. In the case of firm A, the ML prediction of OTE is $\hat{OTE}^1(x_{11}, q_{11}, z_{11}) = 1/2.511 = 0.398$

For a numerical example, reconsider the toy data in Table 1.1. These data have been used to obtain ML and restricted ML (RML) estimates of the unknown parameters in (1.32). The estimates are reported in Table 1.14. Both sets of estimates were obtained under assumptions ML3 and ML4. The RML estimates were obtained by restricting $\lambda \geq 0$. Both sets of estimates have been used to predict levels of OTE. The predictions are reported in Table 1.15. The ML prediction for manager 1 in period 1 is depicted in Fig. 1.18. In this figure, the outputs of firm 1 in period 1 map to point A. The associated predicted frontier output is represented by A*. The dashed line is an estimate of *a function that provides an approximation to* the true frontier depicted earlier in Fig. 1.5.

1.8.3 Productivity Analysis

For purposes of comparison with Sects. 1.6.2 and 1.7.3, this section again focuses on decomposing proper TFPI numbers. Again, both output- and input-oriented decompositions are available. In each case, the precise form of the decomposition depends partly on the SFM.

For an output-oriented example, consider the model defined by (1.32). After some simple algebra, the antilogarithm of this equation can be written as:

$$1 = \exp(\alpha + \lambda t)\left[\prod_{j=1}^{J} z_{jit}^{\delta_j}\right]\left[\prod_{m=1}^{M} x_{mit}^{\beta_m}\prod_{n=1}^{N} q_{nit}^{-\gamma_n}\right]\exp(-u_{it})\exp(v_{it}). \tag{1.38}$$

Multiplying both sides of this equation by $TFP(x_{it}, q_{it})$ yields

$$TFP(x_{it}, q_{it}) = \exp(\alpha + \lambda t)\left[\prod_{j=1}^{J} z_{jit}^{\delta_j}\right]\left[TFP(x_{it}, q_{it})\prod_{m=1}^{M} x_{mit}^{\beta_m}\prod_{n=1}^{N} q_{nit}^{-\gamma_n}\right] \times \exp(-u_{it})\exp(v_{it}) \quad (1.39)$$

A similar equation holds for firm k in period s. Substituting these equations into (1.9) yields

$$\begin{aligned} TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = &\left[\frac{\exp(\lambda t)}{\exp(\lambda s)}\right]\left[\prod_{j=1}^{J}\left(\frac{z_{jit}}{z_{jks}}\right)^{\delta_j}\right] \\ &\times\left[TFPI(x_{ks}, q_{ks}, x_{it}, q_{it})\prod_{m=1}^{M}\left(\frac{x_{mit}}{x_{mks}}\right)^{\beta_m}\prod_{n=1}^{N}\left(\frac{q_{nks}}{q_{nit}}\right)^{\gamma_n}\right] \\ &\times\left[\frac{\exp(-u_{it})}{\exp(-u_{ks})}\right]\left[\frac{\exp(v_{it})}{\exp(v_{ks})}\right] \end{aligned} \quad (1.40)$$

In theory, the presence of statistical noise means we cannot interpret the first three terms in this equation in the same way we interpreted the first three terms in (1.30). However, in practice, the first term would normally be viewed as an output-oriented technology index (OTI), the second term would normally be viewed as an output-oriented environment index (OEI), and the third term would normally be viewed as an output-oriented scale and mix efficiency index (OSMEI). In both theory and practice, the fourth term is an output-oriented technical efficiency index (OTEI), and the last term is a statistical noise index (SNI). Again, the conditions under which these various terms vanish is left as an exercise for the reader.

For an input-oriented example, consider the model defined by (1.34). After some simple algebra, the antilogarithm of this equation can be written as:

$$1 = \exp[\xi(t)]\left[\prod_{m=1}^{M} x_{mit}^{\lambda_m}\prod_{n=1}^{N} q_{nit}^{-\phi_n}\right]\exp(-u_{it})\exp(v_{it}). \quad (1.41)$$

Multiplying both sides of this equation by $TFP(x_{it}, q_{it})$ yields

$$TFP(x_{it}, q_{it}) = \exp[\xi(t)]\left[TFP(x_{it}, q_{it})\prod_{m=1}^{M} x_{mit}^{\lambda_m}\prod_{n=1}^{N} q_{nit}^{-\phi_n}\right] \times \exp(-u_{it})\exp(v_{it}) \quad (1.42)$$

A similar equation holds for firm k in period s. Substituting these equations into (1.9) yields

$$TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) = \left[\frac{\exp[\xi(t)]}{\exp[\xi(s)]}\right] \times \left[TFPI(x_{ks}, q_{ks}, x_{it}, q_{it}) \prod_{m=1}^{M} \left(\frac{x_{mit}}{x_{mks}}\right)^{\lambda_m} \prod_{n=1}^{N} \left(\frac{q_{nks}}{q_{nit}}\right)^{\phi_n}\right] \times \left[\frac{\exp(-u_{it})}{\exp(-u_{ks})}\right]\left[\frac{\exp(v_{it})}{\exp(v_{ks})}\right] \quad (1.43)$$

Again, the presence of statistical noise means we cannot interpret the first two terms in this equation in the same way we interpreted the first and third terms in (1.31). However, in practice, the first term would normally be viewed as an input-oriented

Table 1.16 An output-oriented decomposition of Lowe TFPI numbers using ML[a,b]

Firm	Period	TFPI	OTI	OEI	OTEI	OSMEI	SNI
1	1	1	1	1	1	1	1
2	1	1.786	1	1	1.281	1.394	1.000
3	1	2.37	1	1	2.37	1	1.000
4	1	2.703	1	1	2.314	1.168	1.000
5	1	3.516	1	1	2.474	1.421	1.000
1	2	2.117	0.933	1.196	1.236	1.534	1.000
2	2	3.515	0.933	1.196	2.209	1.426	1.000
3	2	3.513	0.933	1	2.511	1.499	1.000
4	2	2.675	0.933	1.196	0.958	2.503	1.000
5	2	3.159	0.933	1	2.400	1.410	1.000
1	3	3.110	0.871	1	2.344	1.523	1.000
2	3	2.760	0.871	1	1.283	2.469	1.000
3	3	1.879	0.871	1	1.556	1.386	1.000
4	3	3.516	0.871	1	1.578	2.557	1.000
5	3	2	0.871	1.196	1.920	1	1.000
1	4	1.923	0.813	1	1.607	1.472	1.000
2	4	2.032	0.813	1	1.743	1.433	1.000
3	4	1.509	0.813	1.196	1.327	1.170	1.000
4	4	2.852	0.813	1.196	2.441	1.202	1.000
5	4	2.134	0.813	1	1.705	1.539	1.000
1	5	3.150	0.759	1	2.220	1.869	1.000
2	5	2.176	0.759	1.196	2.285	1.050	1.000
3	5	1.991	0.759	1.196	1.599	1.372	1.000
4	5	1.351	0.759	1	1.498	1.188	1.000
5	5	1.758	0.759	1	2.424	0.955	1.000

[a]TFPI = OTI × OEI × OTEI × OSMEI × SNI. Some index numbers may be incoherent at the third decimal place due to rounding (e.g., the product of the OTI, OEI, OTEI, OSMEI and SNI numbers in row 6 is not exactly equal to the TFPI number due to rounding)

[b]Numbers reported to less than three decimal places are exact; see the footnote to Table 1.2 on p. 8

technology index (ITI), and the second term would normally be viewed as an input-oriented scale and mix efficiency index (ISMEI). In both theory and practice, the third term is an input-oriented technical efficiency index (ITEI), and the last term is a statistical noise index (SNI). Again, the conditions under which these terms vanish is left as an exercise for the reader.

For a numerical example, reconsider the toy data in Tables 1.1 and 1.2. The associated Lowe TFPI numbers were reported earlier in Table 1.6. An output-oriented decomposition of these numbers is now reported in Table 1.16. The OTI, OEI and OSMEI numbers in each row were obtained by using the ML estimates in Table 1.14 to evaluate the relevant terms in (1.40). The OTEI numbers were obtained by taking ratios of the ML predictions of OTE reported earlier in Table 1.15. The SNI numbers were obtained as residuals (i.e., SNI $=$ TFPI$/$(OTI$\times$OEI$\times$OTEI$\times$OSMEI)).

1.8.4 Other Models

Other SFMs discussed in this book include various systems of equations. These systems can be used to explain variations in metafrontiers, output supplies and input demands.

1.9 Practical Considerations

This section considers some of the steps involved in conducting a policy-oriented analysis of managerial performance. It also considers government policies that can be used to target the main drivers of performance. In this book, the term ‘government’ refers to a group of people with the authority to control any variables that are not controlled by firm managers.

1.9.1 The Main Steps

Policy-oriented performance analysis involves a number of steps that are best completed in a prescribed order or sequence. The main steps are the following (immediate predecessor steps are in parentheses):

1. Identify the manager(s).
2. Classify the variables that are physically involved in the production process (1).
3. Identify relevant measures of comparative performance (2).
4. Make assumptions about production technologies (2).
5. Assemble relevant data (3).
6. Select functions to represent production possibilities sets (4, 5).
7. Choose an estimation approach (4, 5).

8. Estimate the model and test the model assumptions (6, 7).
9. Check if the main results are robust to the assumptions and choices made in steps 4, 6 and 7 (8).

Researchers with little interest in policy often complete these steps in a different order. For example, academic researchers who are primarily interested in getting their work published often start at Step 7 (i.e., they choose the estimation approach first).

1.9.2 Government Policies

Changes in most measures of managerial performance can be attributed to four main factors: (a) technical progress, (b) environmental change, (c) technical efficiency change, and (c) scale, mix and/or allocative efficiency change. Different government policies affect, and can therefore be used to target, these different components. For example, governments can often increase rates of technical progress by conducting their own R&D, or by directly funding others to conduct R&D. They can often change production environments by, for example, regulating (or failing to regulate) the impact of production processes on the natural environment, and by providing and/or decommissioning different types of public infrastructure. They can often raise levels of technical efficiency by, for example, removing barriers to the adoption of particular technologies, and by providing education and training services to advise managers about the existence and proper use of new technologies. Finally, governments can often raise levels of scale and mix efficiency by changing the variables that drive managerial behaviour. For example, if firms are price-takers in output and input markets, and if managers seek to maximise profits, then governments can often raise levels of scale and mix efficiency by changing relative output and input prices (e.g., by changing minimum wages, interest rates, taxes and/or subsidies).

1.10 Summary and Further Reading

This book is concerned with measuring and explaining changes in managerial performance. The focus is on measures of performance that are useful for policy makers. Most, if not all, of these measures can be viewed as measures of efficiency and/or productivity. The measures of efficiency discussed in this book include measures of technical, scale, mix, revenue, cost, profit and allocative efficiency. Measures of efficiency that are *not* discussed include the measure of marginal cost efficiency discussed by Kutlu and Wang (2018), the measure of environmental efficiency defined by Coelli et al. (2007, Eq. 7), and the measure of irrigation water efficiency defined by Karagiannis et al. (2003, Eq. 2). Most of these other measures can, in fact, be viewed as special cases of the measures discussed in this book.

The measures of productivity discussed in this book include measures of total factor productivity (TFP), multifactor productivity (MFP) and partial factor productivity (PFP). In this book, TFP is defined as a measure of total output quantity divided by a measure of total input quantity. Measures of MFP and PFP can be viewed as measures of TFP in which one or more inputs have been assigned a weight of zero. The definition of TFP used in this book is consistent with concepts and definitions of TFP and/or TFP change that can be found in, for example, Barton and Cooper (1948, p. 123),[5] Jorgenson and Griliches (1967, pp. 249, 250), Christensen and Jorgenson (1970, p. 42), Nadiri (1970, pp. 1138, 1139), Chambers and Pope (1996, p. 1360), Prescott (1998, p. 526) and Good et al. (1999, Sect. 2.1). Elsewhere in the literature, measures of TFP and/or TFP change are often defined in terms of incomes, revenues and/or costs (e.g., Kendrick 1961, p. 10; Foster et al. 2008, p. 400; Lien et al. 2017, p. 253).

This book attributes changes in TFP to four main factors: technical change, environmental change, technical efficiency change, and scale and mix efficiency change. Elsewhere in the literature, it is common to attribute TFP change to (a) technical change only (e.g., Diewert and Morrison 1986, p. 659; Kumbhakar 2002, pp. 469, 471; Orea and Wall 2012, p. 103), (b) a combination of technical change and technical efficiency change (e.g., Nishimizu and Page 1982, pp. 920, 921; Färe et al. 1994, p. 71; Coelli et al. 2003, p. 323), or (c) a combination of technical change and economies of scale (e.g., Kumbhakar et al. 2000, p. 496; Hranaiova and Stefanou 2002, p. 79). In the macroeconomics literature, it is common to equate TFP change with the residuals from regression models (e.g., Olley and Pakes 1996, p. 1287). These alternative approaches are not generally consistent with the way TFP is defined in this book.

Finally, there are several other measures of managerial performance that are not explicitly discussed in this book. These include various measures of corporate social performance. Most of these measures can, in fact, be viewed as measures of TFP. The literature on these measures can be accessed from Siegel and Vitaliano (2007), Chen and Delmas (2011) and Gregory et al. (2016).

References

Barton G, Cooper M (1948) Relation of agricultural production to inputs. Rev Econ Statist 30(2):117–126

Chambers R, Pope R (1996) Aggregate productivity measures. Am J Agric Econ 78(5):1360–1365

Charnes A, Cooper W, Rhodes E (1981) Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Manage Sci 27(6):668–697

Chen CM, Delmas MA (2011) Measuring corporate social performance: an efficiency perspective. Prod Oper Manage 20(6):789–804

Christensen L, Jorgenson D (1970) U.S. real product and real factor input, 1929–1967. Rev Income Wealth 16(1):19–50

Coelli T, Rahman S, Thirtle C (2003) A stochastic frontier approach to total factor productivity measurement in Bangladesh crop agriculture, 1961–92. J Int Dev 15(3):321–333

Coelli T, Lauwers L, van Huylenbroeck G (2007) Environmental efficiency measurement and the materials balance condition. J Prod Anal 28(1–2):3–12

[5] Barton and Cooper (1948) use the term ‘output per unit of input’ instead of TFP.

Cooper W, Seiford L, Thanassoulis E, Zanakis S (2004) DEA and its uses in different countries. Eur J Oper Res 154(2):337–344

Diewert W, Morrison C (1986) Adjusting output and productivity indexes for changes in the terms of trade. Econ J 96(383):659–679

Elteto O, Koves P (1964) On a problem of index number computation relating to international comparison. Statisztikai Sz 42:507–518

Färe R, Grosskopf S (2010) Directional distance functions and slacks-based measures of efficiency. Eur J Oper Res 200(1):320–322

Färe R, Grosskopf S, Norris M, Zhang Z (1994) Productivity growth, technical progress, and efficiency change in industrialized countries. Am Econ Rev 84(1):66–83

Foster L, Haltiwanger J, Syverson C (2008) Reallocation, firm turnover, and efficiency: selection on productivity or profitability? Am Econ Rev 98(1):394–425

Good D, Nadiri M, Sickles R (1999) Index number and factor demand approaches to the estimation of poductivity, vol II. Microeconometrics, Basil Blackwell, Oxford

Gregory A, Whittaker J, Yan X (2016) Corporate social performance, competitive advantage, earnings persistence and firm value. J Bus Finan Acc 43(1&2):3–30

Hranaiova J, Stefanou S (2002) Scale, productivity growth and risk response under uncertainty. Economia Agrar y Recursos Naturales 2(2):73–91

Jorgenson D, Griliches Z (1967) The explanation of productivity change. Rev Econ Stud 34(3):249–283

Karagiannis G, Tzouvelekas T, Xepapadeas A (2003) Measuring irrigation water efficiency with a stochastic production frontier. Env Res Econ 26(1):57–72

Kendrick J (1961) Productivity trends in the United States. Princeton University Press

Kumbhakar S (2002) Risk preference and productivity measurement under output price uncertainty. Empirical Econ 27(3):461–472

Kumbhakar S, Nakamura S, Heshmati A (2000) Estimation of firm-specific technological bias, technical change and total factor productivity growth: a dual approach. Econometric Rev 19(4):162–173

Kutlu L, Wang R (2018) Estimation of cost efficiency without cost data. J Prod Anal 49(2–3):137–151

Lien G, Kumbhakar S, Hardaker J (2017) Accounting for risk in productivity analysis: an application to Norwegian dairy farming. J Prod Anal 47(3):247–257

Nadiri M (1970) Some approaches to the theory and measurement of total factor productivity: a survey. J Econ Lit 8(4):1137–1177

Nishimizu M, Page J (1982) Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia, 1965–1978. Econ J 92(368):920–936

O'Donnell C (2012) Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture. Am J Agric Econ 94(4):873–890

O'Donnell C (2016) Using information about technologies, markets and firm behaviour to decompose a proper productivity index. J Econ 190(2):328–340

O'Donnell C (2017) Estimating total factor productivity change when no price or value-share data are available. Centre for Efficiency and Productivity Analysis Working Papers WP01/2017, University of Queensland

Olley G, Pakes A (1996) The dynamics of productivity in the telecommunications equipment industry. Econometrica 64(6):1263–1297

Orea L, Wall A (2012) Productivity and producer welfare in the presence of production risk. J Agric Econ 63(1):102–118

Prescott E (1998) Lawrence R. Klein Lecture, (1997) Needed: a theory of total factor productivity. Int Econ Rev 39(3):525–551

Siegel D, Vitaliano D (2007) An empirical analysis of the strategic use of corporate social responsibility. J Econ Manage Strategy 16(3):773–792

Szulc B (1964) Indices for multi-regional comparisons. Prezeglad Statystyczny (Statistical Review) 3:239–254