Reference Guide

OnFront® 2

The Professional Tool for Efficiency and Productivity Measurement

Economic Measurement and Quality Corporation

Reference Guide
to
OnFront®
by

Rolf Färe
Professor of Economics

Shawna Grosskopf
Professor of Economics

EMQ AB
Box 2134
SE-220 02 Lund
Sweden

E-mail: emq@emq.com
Internet: www.emq.com
Fax: +46 40 16 45 11

Contents

Theoretical Underpinnings

Introduction and Overview

This Reference Guide is intended to serve as a primer or overview of what we mean by efficiency measurement and productivity measurement—it provides the theoretical underpinnings of the efficiency and productivity measures included in OnFront. It is designed to be self-contained and user friendly; however, it could also be read in conjunction with more detailed discussions such as *Production Frontiers* [1994] by Färe, Grosskopf and Lovell and *Intertemporal Production Frontiers: With Dynamic DEA* [1996] by Färe and Grosskopf.

Simply put, efficiency and productivity measurement tell us about how well a firm (or an agency or even a country) is doing relative to some benchmark. For example, we may wish to know if a particular school district could get more 'bang for the buck' or whether a hospital or government office could provide their current service levels at lower cost. As these examples suggest, we would like our measures to be very flexible—capable of assessing performance even in cases where the usual economic or financial signals like profit or revenues are nonexistent or inappropriate. The models and techniques used in OnFront provide you with this capability.

An important feature of this approach is that performance is judged in a *relative* way. OnFront constructs a benchmark for each individual operation that is based on actual observed achievements in similar operations. We call this benchmark the *best practice frontier*. Section 1 describes in more detail how this best practice frontier is constructed from your data and what it looks like. Included in this section are various characteristics the benchmark may satisfy, including various types of returns to scale and disposability.

After showing how OnFront constructs the benchmark technology, we turn to the various measures of efficiency and productivity that are available in OnFront. There are a number of possibilities, which provide you with a range of choices which vary in terms of the type of data required, and also provide a range of goals or models. For example, if you are looking at pharmacies in Sweden (where the government agreement requires that pharmacies provide quality drugs at lowest cost), it seems reasonable to judge relative performance on that basis. That is, given the level of drugs and services provided at acceptable quality levels, benchmark pharmacies will be those that achieve that goal at lowest cost.

More specifically, we begin with what we call input-saving measures of efficiency. Here benchmark firms are those that produce a given level of goods and services with the fewest resources or lowest cost. Section 2 begins with input-saving technical efficiency, which requires no information on prices, costs or revenues, and proceeds to cost efficiency, which does require information on input prices. Included in this section are various decompositions of the efficiency measures.

Next we turn to output-oriented measures of efficiency. Here the idea is that for a given level of resources used, benchmark firms are those that get the most out of them. Again, we can measure this in terms of input and output quantities, with no information on prices or revenues (technical efficiency), or we can compute revenue efficiency if price information is available.

In Section 4 we turn to productivity measurement, which in our framework, is actually an extension of efficiency measurement to include performance measurement over time.

In Section 5 we include a discussion of our measures of capacity and capacity utilization.

1. Benchmarks: The Best Practice Frontier

In this section we show how OnFront identifies benchmark firms or observations and uses them to construct what we call the *best practice frontier*. This is sometimes also referred to as the reference technology, production frontier or just technology. A particular technology can be constructed for banks, farms, pharmacies, unemployment offices or *any decision-making units* (DMU) that use resources (inputs) to produce outcomes (outputs).

Typically the frontier is constructed from observations of what we call inputs and outputs. There can be any number of inputs (let's say that there are $n = 1,...,N$ different types of inputs), which we will write (individually) as

$$x_n, n = 1,...,N$$

or if we are referring to all the different types of input employed we write this *vector* of inputs as

$$x = (x_1,...,x_N).$$

For each of these inputs, you would have data which would be some nonnegative number. As an example, for farms, inputs would include the data you have on acres of land, hours of farm labor, seed, etc. For a study of unemployment offices inputs might include managers, square meters of office space, etc. As a general principle, these inputs should include all of the resources used by the DMU, and they should be measured as accurately as possible—mistakes in the data (which can create what are called 'outliers') can affect your results.

Turning to outputs, again we allow for many types of output for any DMU; here we assume there are $m = 1,...,M$ of them, denoted individually as

$$y_m, m = 1,...,M$$

or if we are referring to all the different types of output produced we write this *vector* of outputs as

$$y = (y_1,...,y_M).$$

Again, the data for these outputs should be nonnegative numbers. For our farming example, outputs might include bushels of corn, wheat, etc. Outputs from an unemployment

office might include job placements, placements in training programs, etc. Generally, you should include all relevant outputs and services produced or provided by your DMU's.

In constructing the best practice frontier and identifying benchmark firms, we assume that your data includes a number of DMU's or observations; here we assume that there are

$$k = 1,..., K$$

of them. For example, we would say that we have K different observations of unemployment offices. Each observation would include data on all inputs and outputs, i.e.,

$$x^k = (x_{k1},..., x_{kN})$$

and

$$y^k = (y_{k1},..., y_{kM}),$$

where, for example, x_{kn} is the amount of the n^{th} input used by the k^{th} unemployment office, and y_{k1} is the amount of the first service that is produced by k.

Types of Reference Technologies

A reference technology or best practice technology may be expressed in three equivalent ways, including

- an *Input Requirement Set* $L(y)$ that shows all the combinations of inputs that can be used to produce the output vector y,
- an *Output Possibility Set* $P(x)$ which shows all the combinations of outputs that can be produced by the input vector x,
- a *Graph* GR which shows the combinations of inputs x and outputs y that are technically feasible.

Again, these are all equivalent, but can be used to illustrate different aspects of production.

Numerical Example: The Input Requirement Set

We begin with the *Input Requirement Set*, which is particularly useful in illustrating input substitution possibilities. We construct input requirement sets from the data or observations of inputs and outputs, $(x^k, y^k), k = 1,..., K$. This technology can be constructed from your data using what we call activity analysis. It can be written in the following way

$$
\begin{aligned}
L(y \mid C, S) = \{(x_1,..., x_N) : & \qquad (1)\\
& \sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1,..., M,\\
& \sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1,..., N,\\
& z_k \geqq 0, k = 1,..., K\},
\end{aligned}
$$

where the $z_k, k = 1,..., K$ are what are referred to as the intensity variables. We will show below what role they play in constructing the reference technology.

Before we discuss the role of the C and S in (1), let's construct a reference technology using some real numbers, based on the table below.

Firm (DMU)	Input 1 x_1	Input 2 x_2	Output y
1	1	2	1
2	2	1	1
3	2	2	1

Table 1: Data for a Best Practice Technology

The table shows that there are three firms (DMU's) that use two inputs x_1 and x_2 to produce a single output y. The amount of input one used by firm two is $x_{21} = 2$, for example.

In order to illustrate how the best practice technology can be constructed from the data in Table 1 we use both a diagram and explicit equations based on the data. In equation form, using the data from the table and substituting into the equations in (1), we get

$$
\begin{aligned}
L(y \mid C, S) = \{(x_1, x_2) : & \qquad (2)\\
& z_1 1 + z_2 1 + z_3 1 \geqq y, \qquad \text{(a)}\\
& z_1 1 + z_2 2 + z_3 2 \leqq x_1, \qquad \text{(b)}\\
& z_1 2 + z_2 1 + z_3 2 \leqq x_2, \qquad \text{(c)}\\
& z_1 \geqq 0, z_2 \geqq 0, z_3 \geqq 0\},
\end{aligned}
$$

and the diagram is given below.

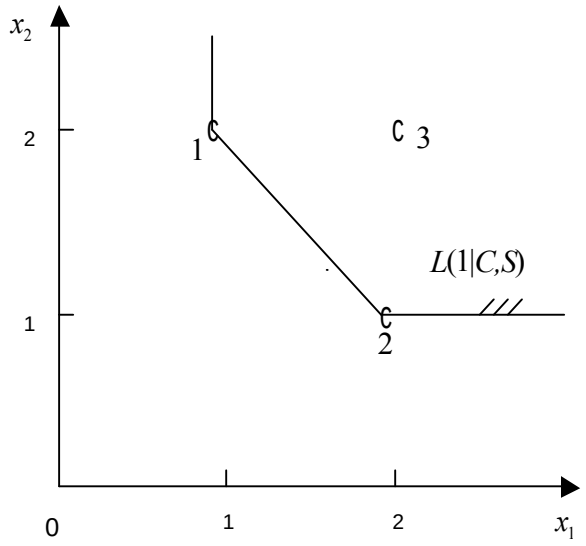

Figure 1: An Input Requirement Set

In the figure the three firms are labeled as 1, 2 and 3, where, for example, $x^1 = (1,2)$ and $x^2 = (2,1)$. Notice that they all produce the same amount of output $y = 1$. If we let the intensity variables take the values $z_1 = 1,\ z_2 = z_3 = 0,$ we can see from the equations above that we will get back $x^1 = (1,2)$ as a feasible member of the input requirement set $L(1 \mid C,S)$. If we instead set the intensity variables to the values $z_2 = 1,\ z_1 = z_3 = 0,$ we would get back the observed values of x for firm 2, i.e., the point labeled 2 in the figure. We can also construct 'hypothetical' members (including benchmark members) by varying the values of the intensity variables: for example if we let $z_1 = z_2 = 1/2$ and $z_3 = 0$, we will end up with a point or hypothetical observation that is halfway between 1 and 2 in the figure. In fact, all convex combinations of DMU's 1,2, and 3 are feasible (hypothetical) members of the input requirement set. In addition, due to the inequalities (rather than strict equalities) in (b) and (c) in (2), all points northeast of the line segment between 1 and 2 are also feasible (but hypothetical) members of the input requirement set. From the figure it is easy to see the benchmarks or frontier of this input requirement set: they are the points that form the 'lower' boundary of the set, i.e., they represent real or hypothetical firms that use the fewest possible inputs to produce the same outputs.

Disposability Properties of $L(y)$

Next we turn to a brief discussion of what we call the disposability properties of the input requirement set. The inequalities (b) and (c) in (2), or more generally the $n = 1,..., N$ inequalities in (1), model what we call *Strong Disposability of Inputs*, i.e.,

$$x \geqq \hat{x} \in L(y \mid C,S) \text{ implies that } x \in L(y \mid C,S). \tag{3}$$

In words this says that if inputs are either held the same or increased, then output will not decrease. Strong disposability of inputs means that an increase in inputs cannot decrease, i.e., 'congest' output. What we mean by congestion is that there is 'too much' input. Examples where too much input can actually obstruct or reduce output include traffic congestion, where too many cars on the road at one time leads to slower traffic.

To allow for the possibility of congestion in our models we must use a different assumption about disposability of inputs which we call *Weak Disposability of Inputs*, i.e.,

$$x \in L(y \mid C,W) \text{ and } l \geqq 1 \text{ imply } lx \in L(y \mid C,W). \tag{4}$$

This disposability assumption states that proportional increases in inputs do not decrease outputs. Note that we use a 'W' rather than an 'S' in the notation for the input requirement set to distinguish weak from strong disposability. We can incorporate weak disposability into our activity analysis model (1) by changing the inequalities in the N input constraints to strict equalities, i.e.,

$$\begin{aligned} L(y \mid C,W) = \{(x_1,...,x_N): & \\ & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1,...,M, \\ & \sum_{k=1}^{K} z_k x_{kn} = x_n, n = 1,...,N, \\ & z_k \geqq 0, k = 1,...,K\}. \end{aligned} \tag{5}$$

Model (1) and (5) give the two extremes with respect to disposability of inputs. In the first model, all of the inputs are strongly or freely disposable, while in the second none of the inputs are strongly disposable. You can construct intermediate cases by restricting only some of the inputs to be strongly disposable, for example. Generally speaking, if you feel that some inputs may be disposed of without any resource cost, then you would use inequalities for those constraints. If you feel that other inputs may cause congestion, then you would use equalities for those constraints. For example

$$\begin{aligned} & \sum_{k=1}^{K} z_n x_{kn} = x_n, n = 1,...,\hat{N} \\ & \sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = \hat{N}+1,...,N, \end{aligned} \tag{6}$$

where the first $n = 1,...,\hat{N}$ inputs may cause congestion, while the $\hat{N}+1,...,N$ remaining inputs can be disposed of at no cost.

Modeling Returns to Scale

Now we turn to what we mean by 'C' in $L(y \mid C,S)$. It refers to the type of returns to scale we have specified for the reference technology: 'C' refers to *Constant Returns to Scale*, which is defined as

$$L(qy \mid C,S) = qL(y \mid C,S), q > 0. \tag{7}$$

Under constant returns to scale, proportional changes in outputs require proportional changes in inputs. For example, if you plan to double your outputs, under constant returns to scale you must double your employment of all inputs. The particular input requirement set described in (1) satisfies constant returns to scale due to the restriction on the intensity variables: namely, they are only restricted to be nonnegative, $z_k \geqq 0, k = 1,...,K$. In other words, they may be scaled up and down by any positive scalar q. In turn, they serve to scale the data up and down by that amount.

From economic theory we know that an organization or firm that operates under constant returns to scale earns zero profit (although their revenues cover all of their costs, including opportunity costs). This is one of the reasons one may wish to consider allowing for alternative assumptions concerning returns to scale. One alternative is to allow for a reference technology that exhibits *Nonincreasing Returns to Scale* (N), which is defined as

$$L(qy \mid N,S) \subseteq qL(y \mid N,S), 0 \leqq q \leqq 1, \tag{8}$$

where $\subseteq$ means included in. In this case if you wish to scale up your outputs, then you would need to scale up your inputs by a larger amount. This property can be incorporated into our model by changing the restriction on the intensity variables z to add up to no more than one, i.e.,

$$\begin{aligned} L(y \mid N,S) = \{(x_1,...,x_N): & \\ & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1,...,M, \\ & \sum_{k=1}^{k} z_k x_{kn} \leqq x_n, n = 1,...,N, \\ & \sum_{k=1}^{K} z_k \leqq 1, z_k \geqq 0, k = 1,...,K\}. \end{aligned} \tag{9}$$

To illustrate how the restrictions on the intensity variables result in constant returns to scale and nonincreasing returns to scale technologies, we turn to a numerical example. Suppose we are given the following data in Table 2.

Firm (DMU)	Input x	Output y
1	1	2
2	2	3

Table 2: Data to Illustrate Returns to Scale

In this case we have two firms, each using one input to produce one output. The two reference technologies $L(y \mid C,S)$ and $L(y \mid N,S)$ can be written in equation form as

$$\begin{aligned} L(y \mid C,S) = \{x : & \\ & z_1 2 + z_2 3 \geqq y, \\ & z_1 1 + z_2 3 \leqq x, \\ & z_1 \geqq 0, z_2 \geqq 0\}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} L(y \mid N,S) = \{x : & \\ & z_1 2 + z_2 3 \geqq y, \\ & z_1 1 + z_2 2 \leqq x, \\ & z_1 + z_2 \leqq 1, \\ & z_1 \geqq 0, z_2 \geqq 0\}, \end{aligned} \tag{11}$$

We can also illustrate this in a diagram, this time in terms of the graph (GR), which shows the relationship between inputs and outputs and therefore is useful in illustrating returns to scale.

First of all our two observations are labeled 1 and 2, respectively, in the diagram. The constant returns to scale (C) technology is bounded by the x-axis and the ray through point a. The nonincreasing returns to scale technology is bounded by the line segments $0a$, $0b$, the horizontal extension from b and the x-axis. The nonincreasing returns technology does not allow for outward scaling (like the ray extending beyond point a of the constant returns technology); it does allow for the convex combinations of existing observations and the origin, i.e., the radial contractions. The extension beyond point b on the nonincreasing returns boundary follows from the inequality on the input constraint, i.e., it is due to the fact that we are allowing for strong disposability of inputs in (11). From the diagram we can also see that

$$L(y \mid N,S) \subseteq L(y \mid C,S), \tag{12}$$

which is true in general.

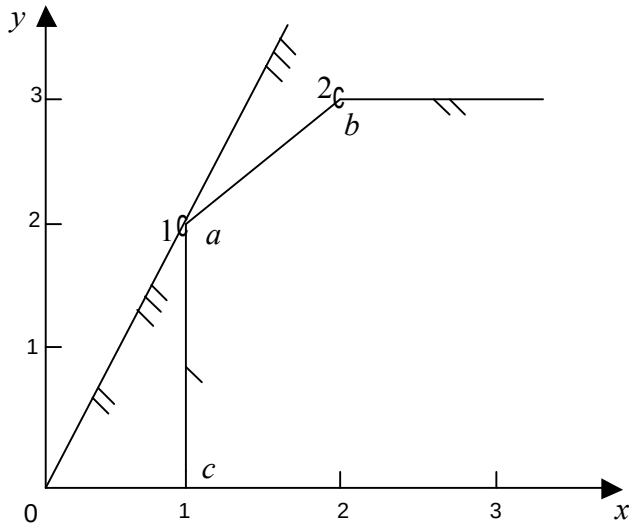

Figure 2: Constant and Nonincreasing Returns to Scale

If we restrict the intensity variables to sum exactly to one, rather than to less than or equal to one, we obtain what we refer to as a *Variable Returns to Scale* reference technology. Specifically,

$$L(y \mid V,S) = \{(x_1,...,x_N): \quad (13)$$
$$\sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1,...,M,$$
$$\sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1,...,N,$$
$$\sum_{k=1}^{K} z_k = 1, z_k \geqq 0, k = 1,...,K\}.$$

Here the 'V' refers to variable returns to scale, which is obtained through the restriction $\sum_{k=1}^{K} z_k = 1$. In Figure 2, the variable returns to scale technology is bounded by the *x*-axis starting at $x = 1$, and the line segments *ca*, *ab* and the horizontal line emanating from *b*. From the diagram and inspection of the restrictions on the *z* variables, we can now summarize

$$L(y \mid V,S) \subseteq L(y \mid N,S) \subseteq L(y \mid C,S). \quad (14)$$

We will make use of this relationship when we show how to measure scale efficiency, or deviations from constant returns to scale.

Modeling Technology with the Output Set

We mentioned earlier that the input requirement set, the output possibilities set and the graph are all equivalent ways of describing the technology. Here we focus on the output set, which gives us the set of all combinations of outputs that are producible from a given set of inputs. For the case of constant returns to scale (C) and strong disposability of *outputs* (S), we have the following

$$P(x \mid C,S) = \{(y_1,...,y_M): \quad (15)$$
$$\sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1,...,M,$$
$$\sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1,...,N,$$
$$z_k \geqq 0, k = 1,...,K\}.$$

As before 'C' stands for constant returns to scale, however, 'S' now stands for strong disposability of *outputs*. Note that in (15) we also have strong disposability of inputs (due to the inequality on the input constraints), however, here we will focus on disposability of outputs. *Strong Disposability of Outputs* is defined as

$$y \leqq \hat{y} \in P(x \mid C,S) \text{ implies that } y \in P(x \mid C,S). \quad (16)$$

In words this says that outputs can be disposed of without cost, i.e., you can 'freely' dispose of outputs. This free disposability follows from the inequalities on the output restrictions in (15).

Although free or strong disposability is the assumption we typically make about disposability of outputs, it cannot model some interesting cases where disposal is in fact costly. For example, coal fired electric utilities produce SO_2 as well as electricity. Although electricity is likely to be freely disposable, under current environmental regulations, SO_2 is not. To allow for this type of case, we introduce the idea of *Weak Disposability of Outputs* as

$$\text{If } y \in P(x \mid C,W) \text{ and } 0 \leqq q \leqq 1 \text{ then } qy \in P(x \mid C,W). \quad (17)$$

This says that proportional reductions of all outputs are feasible; however, it does not necessarily follow that reductions in individual outputs are feasible. The proportional reduction implies that to reduce undesirable outputs like SO_2 is costly in the sense that it uses resources which otherwise could have been used to maintain or increase desirable outputs like electricity.

We can model weak disposability of outputs by replacing the inequalities for outputs in (15) by equalities,

$$P(x \mid C,W) = \{(y_1,...,y_M): \quad (18)$$

$$\sum_{k=1}^{K} z_k y_{km} = y_m, m = 1,...,M,$$
$$\sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1,...,N,$$
$$z_k \geqq 0, \quad k = 1,...,K\}.$$

To illustrate the difference between the two disposability ideas, we provide another numerical example. The data in Table 3 is for two firms which use one input to produce two outputs.

Firm (DMU)	Input x	Output 1 y_1	Output 2 y_2
1	1	3	1
2	1	1	2

Table 3: Data for a Best Practice Technology

The two firms in Figure 3 are represented by the points labeled *b* and *c*. The output set satisfying weak disposability of outputs is bounded by 0*bc*0 and the output set satisfying strong disposability of outputs is bounded by 0*abcd*0. This illustrates the following general relationship of inclusion,

$$P(x \mid C,W) \subseteq P(x \mid C,S). \tag{19}$$

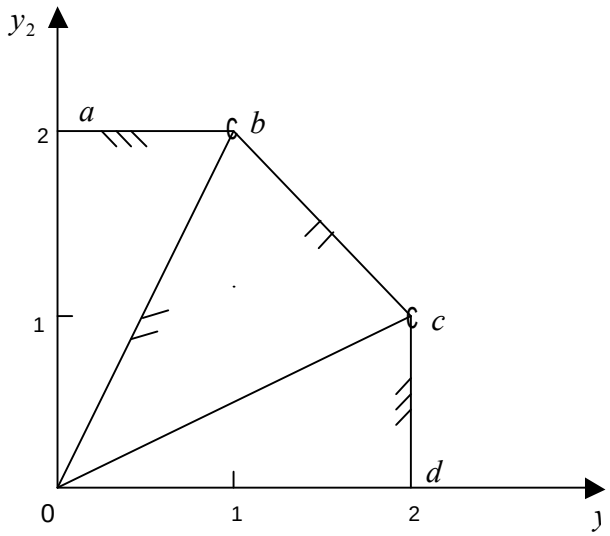

Figure 3: Output Disposability

To summarize, if you wish to model the case where undesirable outputs, like waste, are produced simultaneously with desirable outputs, you might want to include both weak and strong disposability. In this case you would require that the 'bads' satisfy weak disposability and the 'goods' satisfy strong disposability, i.e.,

$$\sum_{k=1}^{K} z_k y_{km} = y_m, m = 1,...,\hat{M}, \qquad (20)$$
$$\sum_{k=1}^{k} z_k y_{km} \geqq y_m, m = \hat{M}+1,...,M,$$

where the first $m = 1,...,\hat{M}$ outputs may cause congestion, while the $\hat{M}+1,...,M$ remaining outputs can be disposed of at no cost.

Next we turn to modeling returns to scale with the output set. This is exactly analogous to the way we modeled returns to scale using the input requirement set, namely by changing the restrictions on the z or intensity variables. This is summarized in the table below.

Technology	Returns To Scale	Restriction On z's	
$P(x \mid C,S)$	C	$z_k \geqq 0,$	$k = 1,...,K$
$P(x \mid N,S)$	N	$\sum_{k=1}^{K} z_k \leqq 1, z_k \geqq 0,$	$k = 1,,...,K$
$P(x \mid V,S)$	V	$\sum_{k=1}^{K} z_k = 1, z_k \geqq 0,$	$k = 1,...,K$

Returns to Scale and the Output Set

The restrictions for the various types of returns to scale are exactly the same as those required for the input sets. This follows from the equivalence of the input and output sets.

2. Direct Input-Saving Efficiency Measures

In the previous section we described how the reference technology is constructed from your data. Now we turn to the problem of measuring how well individual firms or DMU's are doing given the best practice available. Economists typically think of firms as 'optimizing', i.e., they have a goal and they make production choices to do the best they can in achieving that goal given the technology constraints they face. The textbook example of a firm optimization problem is profit maximization. All of the efficiency measures we discuss here are essentially consistent with that goal, although we will focus on more narrow aspects of that problem.

For example, the efficiency measures discussed in this section are all based on part of the profit maximization problem, namely the minimization of costs or resource use. For a given level of production, 'efficiency' requires that firms use the fewest possible resources, i.e., they are saving or reducing inputs (costs) as much as possible. As a consequence we refer to these as input-saving or input-oriented measures of efficiency. These are in contrast to output-oriented efficiency measures (discussed later) where the idea is to produce as much output or revenue as

possible from a given level of resources. Under certain conditions input- and output-oriented efficiency measures will give the same result, but in general that is not the case.

Among the input efficiency measures we distinguish between those that require data on prices and those that do not require price data. The measures which require price data might be referred to as economic measures and those that don't as technical measures of efficiency. We begin with the 'simpler' measures, namely technical efficiency.

Technical efficiency is easiest to understand by way of a diagram. Since inputs are our focus here, we use an input requirement set as our reference technology. Figure 4 is a replica of Figure 1, where we have 3 firms or DMU's, labeled 1, 2, and 3. Notice that firms 1 and 2 are on the boundary of the input requirement set, while firm 3 is inside the boundary. In this example it is clear that 1 and 2 are benchmark firms, whereas firm 3 is not. We would say that firms 1 and 2 are technically efficiency and firm 3 is technically inefficient. The problem is to find a way to summarize how inefficient firm 3 is relative to the benchmark firms. The reference technology allows us to compare firm 3 to a (in this case hypothetical) benchmark firm that has exactly the same mix of inputs as firm 3, namely the hypothetical firm at point *b*. Point *b* is on the lower boundary of the input requirement set $L(1 \mid C, S)$, i.e., it is on the best practice frontier.

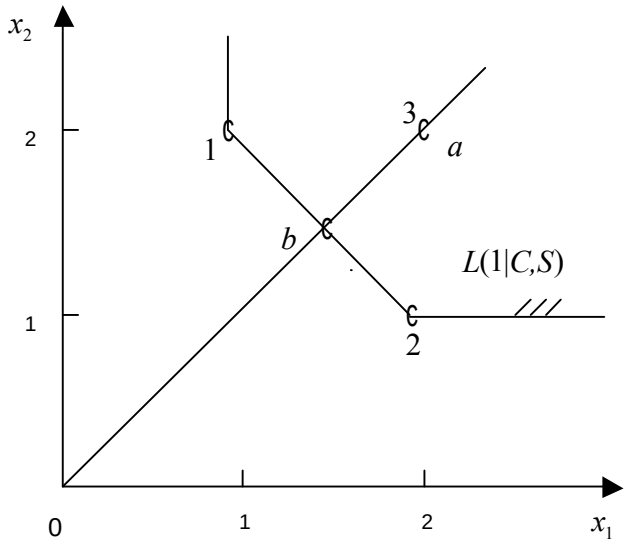

Figure 4: The Input Measure of Technical Efficiency

In choosing point b as our comparison or benchmark, we are choosing to measure inefficiency in a proportional way, i.e., the mix of inputs (the ratio of x_1 / x_2) is the same at point *a* and point *b*. In principle, firm 3 should be able to scale down its input use to point *b* and still produce the same amount of output. We measure technical efficiency as the greatest proportion that inputs could be reduced and still produce the same output. Alternatively, it is the ratio of (the size of) minimal feasible input usage to (the size of) current input usage. In our figure that is

$$0a / 0b.$$

This measure is usually referred to as *the Farrell Input-Saving Measure of Technical Efficiency*, and it is formally defined as

$$F_i(y,x \mid C,S) = \min\{l : lx \in L(y \mid C,S)\}. \tag{21}$$

To show how this could be calculated, we can use the data from Table 1 (which was used to construct the input requirement set in the diagram above). We will set up the problem for firm 3, which uses two inputs $x_1 = 2$ and $x_2 = 2$ to produce one unit of output $y = 1$. For firm 3 we have

$$\begin{aligned} F_i(1,2,2 \mid C,S) = \min\ & l \\ \text{s.t.}\quad & z_1 1 + z_2 1 + z_3 1 \geqq 1, \\ & z_1 1 + z_2 2 + z_3 2 \leqq l\,2, \\ & z_1 2 + z_2 1 + z_3 2 \leqq l\,2, \\ & z_1 \geqq 0, z_2 \geqq 0, z_3 \geqq 0, \end{aligned} \tag{22}$$

which has a solution value of

$$F_i(1,2,2 \mid C,S) = 0.75,$$

which says that firm 3's inputs could be scaled back by multiplying them by 0.75, which is equivalent to a 25% reduction. That would give us input usage of $x_1 = x_2 = 1.5,$ which is exactly the amount at point *b*, our benchmark firm.

If instead we measure the efficiency of firm 1, we would get

$$F_i(1,1,2 \mid C,S) = 1.00,$$

i.e., we cannot reduce firm 1's inputs and still produce the same output, since it is a benchmark firm. In general, we have the following result that

$$0 < F_i(y,x \mid C,S) \leqq 1, \tag{23}$$

and we say that a firm *k* is technically input efficient if

$$F_i(y^k, x^k \mid C,S) = 1, \tag{24}$$

and inefficient if the value is less than one.

Input Slack

Intuitively, firms are efficient if they are on the boundary of the best practice frontier, i.e., if they are benchmark firms. In some cases, however, these benchmark firms may not be using the fewest possible inputs to produce their output. To see what we mean, suppose we add another firm to our data set with the inputs and outputs listed in Table 4.

Firm (DMU)	Input 1 x_1	Input 2 x_2	Output y
4	3	1	1

Table 4: Added Data

If we compute the efficiency for this additional DMU 4 relative to the reference technology constructed from Tables 1 and 4, it follows that

$$F_i(1,3,1 \mid C,S) = 1.00.$$

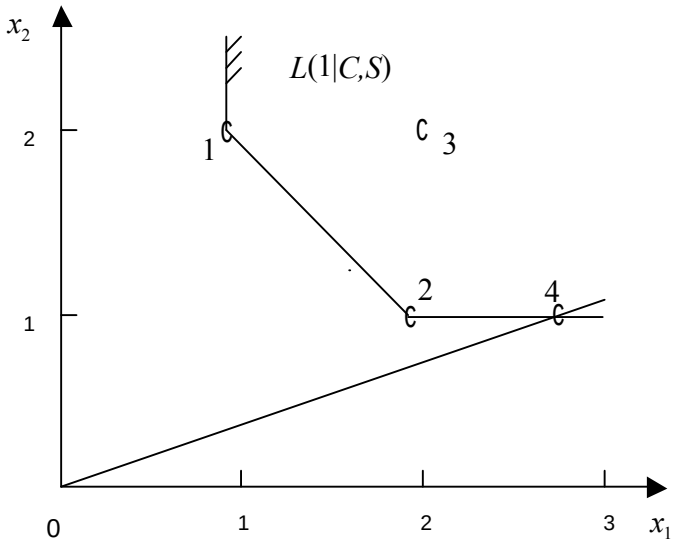

Figure 5: Input Slack

However, if we look at our diagram with our additional data point added in (see Figure 5), we can see that firm 4 could actually produce 1 unit of y with less x_1, i.e., it could reduce x_1 by one unit (from 3 to 2). This extra input is referred to as input slack (although technically what we have in this case is a surplus of input x_1). In general, we say that there is *Slack* in input x_n for firm k' if

$$\sum_{k=1}^{K} z_k x_{kn} < x_{k'n} F_i(y^{k'}, x^{k'} \mid C,S), \tag{25}$$

is true for some solution value for z_k, $k = 1,...,K$.

Decomposing Technical Efficiency

Next we show how our input technical efficiency measure can be decomposed into components: scale efficiency, a measure of congestion and a residual technical efficiency component.

The Scale Efficiency Component

We have already shown that by changing the restriction on the intensity variables, z_k, $k = 1,...,K,$ you can change the returns to scale properties of the reference technologies. Recall that if $\sum_{k=1}^{K} z_k \leqq 1$ we have *Nonincreasing Returns to Scale* (N) and if $\sum_{k=1}^{K} z_k = 1$ we have *Variable Returns to Scale* (V). In order to define scale efficiency we must first introduce an input measure of efficiency which is measured relative to a variable returns technology, $L(y \mid V,S)$, i.e.,

$$F_i(y,x \mid V,S) = \min\{l : lx \in L(y \mid V,S)\}. \tag{26}$$

If we use the data from Table 2 along with the corresponding Figure 2, we find that for DMU 1

$$F_i(2,1 \mid V,S) = 1.0,$$

and that for DMU 2,

$$F_i(3,2 \mid V,S) = 1.0.$$

This is easy to see from Figure 2. Both DMU 1 and 2 are on the boundary of the V technology (bounded by *cab*, the extension from *b* and the *x*-axis from *c* outward).

It we write out the equations for the linear programming problem for DMU 2 we get

$$\begin{aligned} F_i(3,2 \mid V,S) = \min\ & l \\ \text{s.t.}\quad & z_1 2 + z_2 3 \geqq 3, \\ & z_1 1 + z_2 2 \leqq l\,2, \\ & z_1 + z_2 = 1, \\ & z_1 \geqq 0,\ z_2 \geqq 0. \end{aligned} \tag{27}$$

The solution value of 1 is consistent with values of $z_1 = 0$ and $z_2 = 1$, which satisfies our restriction on the *z* variables.

If we compute the efficiency scores for the same data relative to a constant returns to scale technology (where the *z*'s are only restricted to be nonnegative),we find that for DMU 1

$$F_i(2,1 \mid C,S) = 1.0,$$

and that for DMU 2,

$$F_i(3,2 \mid C,S) = 0.75.$$

In Figure 2, DMU 2 is at point *b*, which is not on the frontier of the constant returns to scale technology; for an output level of 3, a point on the C frontier would use only 1.5 instead of 2 units of input. In Figure 2, where we have the graph of technology, input-saving means moving due West in the figure.

Thus DMU 1 is efficient relative to both reference technologies, and DMU 2 is efficient only relative to the V technology. Since it is inefficient relative to the constant returns to scale technology, that means that it deviates from constant returns to scale. This deviation is the intuition behind our measure of *Input Scale Efficiency*, which is defined as

$$S_i(y,x \mid S) = F_i(y,x \mid C,S) / F_i(y,x \mid V,S). \tag{28}$$

A DMU is scale efficient if $S_i(y,x \mid S) = 1$, i.e., if $F_i(y,x \mid C,S) = F_i(y,x \mid V,S)$. This is true for DMU 1, while DMU 2 is scale inefficient with a score

$$S_i(3,2 \mid S) = 0.75.$$

The scale efficiency measure tells us whether a DMU is operating at a point where constant returns prevail, however, if a DMU is scale inefficient we can't tell from the scale efficiency score whether that is due to the fact that they are operating at increasing or decreasing returns to scale. To do so, we need a little more information. If we compute input inefficiency relative to a nonincreasing returns to scale technology, $L(y \mid N,S)$, i.e.,

$$F_i(y,x \mid N,S) = \min\{l : lx \in L(y \mid N,S)\}, \tag{29}$$

we can use this to identify the returns to scale. In our example, $F_i(3,2 \mid N,S) = 1$ for DMU 2, which is consistent with the following relationship.

$$F_i(y,x \mid V,S) \leqq F_i(y,x \mid N,S) \leqq F_i(y,x \mid C,S), \tag{30}$$

which for DMU 2 is $1 \leqq 1 \leqq .75$. Since, in this case, $F_i(y,x \mid V,S) = F_i(y,x \mid N,S)$ we know that DMU 2 is operating at a point of decreasing returns to scale, which is confirmed by referring to Figure 2. In general, we can say that

$$\text{if } S_i(y,x \mid S) < 1, \quad \text{then scale inefficiency is due to:} \tag{31}$$
$$\text{increasing returns to scale if } F_i(y,x \mid N,S) = F_i(y,x \mid C,S),$$
$$\text{decreasing returns to scale if } F_i(y,x \mid N,S) > F_i(y,x \mid C,S).$$

Thus by computing the three efficiency measures,

$$F_i(y,x \mid V,S), F_i(y,x \mid N,S), F_i(y,x \mid C,S),$$

we can determine if a DMU is scale efficient. If it is not scale efficient, we can determine whether scale efficiency is due to operation at decreasing or increasing returns to scale. In addition we obtain the following decomposition of our original input-saving measure of technical efficiency (21)

$$F_i(y,x \mid C,S) = S_i(y,x \mid S) \cdot F_i(y,x \mid V,S). \tag{32}$$

This decomposition shows that a scale efficiency measure can be extracted from $F_i(y,x \mid C,S)$. Next we show how to also identify a congestion component.

The Congestion Component

In order to isolate congestion, we need to compute the following input-saving measure

$$F_i(y,x \mid V,W) = \min\{l x \in L(y \mid V,W)\}, \tag{33}$$

where the technology satisfies variable returns to scale (V) and weak disposability of inputs (W). We can now define the *Input Congestion Measure* as

$$CN_i(y,x \mid V) = F_i(y,x \mid V,S) / F_i(y,x \mid V,W). \tag{34}$$

We say that an observation *k* is congestion free if $CN_i(y,x \mid V) = 1,$ and that it is congesting if the measure is less than one. Combining (32) and (34) we can now summarize the following decomposition of our original input measure of technical efficiency from (21), i.e.,

$$F_i(y,x \mid C,S) = S_i(y,x \mid S) \cdot CN_i(y,x \mid V) \cdot F_i(y,x \mid V,W). \tag{35}$$

The first component measures deviations from constant returns to scale, the second captures deviations from strong disposability of inputs and the third is a measure of input technical efficiency measured relative to a variable returns to scale technology with weak disposability.

The next section turns to ‘economic’ or price related measures of input efficiency.

Input-Saving Measures of Efficiency with Prices

We call these price related measures 'economic', since economists are typically interested in prices, costs, etc. For the input-saving economic measures of efficiency, we assume that input prices $w^k = (w_{k1},...,w_{kN})$ are known for each firm or DMU. If we know both input prices and input quantities, we can also compute *Total Cost* (C^k) for each DMU, $k = 1,...,K$

$$C^k = \sum_{n=1}^{N} w_{kn} x_{kn} = w^k x^k. \tag{36}$$

As an example, suppose that the four DMU's from Table 1 and 4 all face the same input prices $w^k = (1,2)$, $k = 1,...,4$, then we have the following information

Firm (DMU)	Input 1 x_1	Input 2 x_2	Output y	Total Cost
1	1	2	1	5
2	2	1	1	4
3	2	2	1	6
4	3	1	1	5

Table 5: Total Cost

Of course, these costs may not be minimal cost for the individual DMU's when we take into account the best practice frontier. This is easy to see if we plot the data, see Figure 6, which also allows us to illustrate the famous Farrell decomposition of cost efficiency.

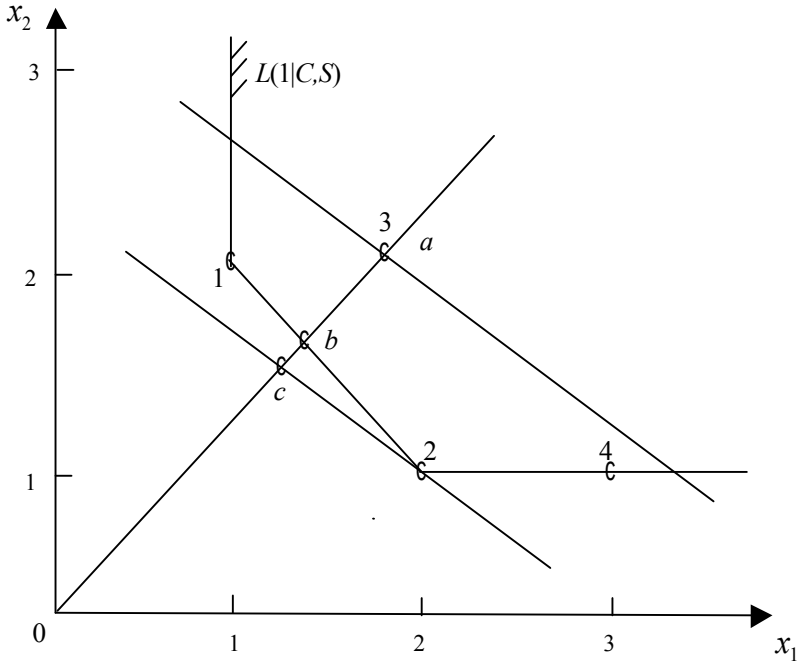

Figure 6: The Input-Output Farrell Decomposition

In Figure 6, the reference technology $L(1 \mid C,S)$ is constructed from the data in Table 5. If you focus on DMU 3 as an example, you see that it uses two units of each input to produce one unit of output. Given our input prices $w_{k1} = 1$ and $w_{k2} = 2$, DMU 3 has the following total cost

$$C^3 = 6 = 1 \cdot 2 + 2 \cdot 2.$$

From the figure you see that DMU 3 is technically inefficient, and from (21) this technical inefficiency equals

$$F_i(1,2,2 \mid C,S) = 0.75.$$

This technical inefficiency can be given a cost interpretation since we know input prices. If DMU were technically efficient, it would be producing at *b* and its cost would be

$$1 \cdot 1.5 + 2 \cdot 1.5 = 4.5,$$

which we can compare to its actual total cost of \$6 to get

$$4.5 / 6 = 0.75,$$

which corresponds to DMU 3's measure of technical efficiency, i.e., DMU 3 could reduce costs by 25% if they eliminated technical inefficiency.

That isn't the end of the story, because in terms of cost, DMU 3 could do even better by operating with the same inputs as DMU 2. This would require changing its input mix, but that makes sense given the relative prices. We can compute this *Minimum Cost* for DMU 3 as the solution to

$$\begin{aligned} C(1,1,2 \mid C,S) = \min\ & 1x_1 + 2x_2 \qquad (37) \\ \text{s.t.}\quad & z_1 1 + z_2 1 + z_3 1 + z_4 1 \geqq 1, \\ & z_1 1 + z_2 2 + z_3 2 + z_4 3 \leqq x_1, \\ & z_1 2 + z_2 1 + z_3 2 + z_4 1 \leqq x_2, \\ & z_1 \geqq 0, z_2 \geqq 0, z_3 \geqq 0, z_4 \geqq 0, \end{aligned}$$

which for our DMU 3 is

$$C(1,1,2 \mid C,S) = 4.$$

Although the programming problem in (37) looks a lot like those used to compute technical efficiency, there are some differences. First the objective (the first line) is different; here we are explicitly seeking to minimize the cost of inputs. Another difference is that the input constraints now have the input quantities of the DMU being evaluated as unknowns or variables rather than given quantities. That is what allows for the change in input mix. In the technical efficiency

measures, the input *mix* (which is defined as $x/\|x\|$) was not allowed to change; inputs were changed by scaling all of them down by the same proportion (l). Figure 6 illustrates: technical efficiency scales along a ray, whereas identification of minimal cost can result in changes in input mix.

The general statement of the *Cost Minimization Problem* is

$$C(y,w\,|\,C,S)=\min\left\{\sum_{n=1}^{N} w_n x_n : (x_1,...,x_N)\in L(y\,|\,C,S)\right\}. \tag{38}$$

Although we have defined minimum cost relative to $L(y\,|\,C,S)$ here, you could minimize cost relative to other technologies. Our definition here allows us to have a decomposition that is analogous to our decomposition of $F_i(y,x\,|\,C,S)$.

We begin by defining *Cost Efficiency* as the ratio of minimum to total (observed) cost, i.e.,

$$O_i(y,x,w\,|\,C,S)=C(y,w\,|\,C,S)/wx,$$

which for DMU 3 is equivalent to $0a/0c$ in Figure 6, or in terms of the costs

$$(1\cdot 2+2\cdot 1)/(1\cdot 2+2\cdot 2)=4/6.$$

Notice that for DMU 3 we have identified cost efficiency as $0a/0c$ and technical efficiency as $0a/0b$. The residual between these two is generally referred to as *Input Allocative Efficiency* and can be computed as

$$A_i(y,x,w\,|\,C,S)=O_i(y,x,w\,|\,C,S)/F_i(y,x\,|\,C,S), \tag{39}$$

which for DMU 3 is $0b/0c$. If we rearrange the expression for allocative efficiency we arrive at what is known as the *Farrell Decomposition of Input Efficiency*,

$$O_i(y,x,w\,|\,C,S)=A_i(y,x,w\,|\,C,S)\cdot F_i(y,x\,|\,C,S), \tag{40}$$

In terms of Figure 6, for DMU 3 we have

$$0c/0a=(0c/0b)(0b/0a).$$

If we combine the Farrell decomposition (40) with our decomposition of the technical efficiency measure (32), we have

$$O_i(y,x,w\,|\,C,S)=A_i(y,x,w\,|\,C,S)\cdot S_i(y,x\,|\,S)\cdot CN_i(y,x\,|\,V)\cdot F_i(y,x\,|\,V,W). \tag{41}$$

This grand decomposition of our economic and technical input-saving measures of efficiency shows that input cost efficiency can be broken down into four different components: one economic and three technical. The economic component is allocative efficiency, the technical components include scale efficiency, congestion and technical efficiency, where the last is measured relative to the $L(y \mid V, W)$ technology.

Subvector Efficiency

So far in our efficiency measures, inputs have been treated symmetrically. For example, the technical efficiency measures scales each input with the same (l) factor and the economic measure minimizes the cost of all inputs. At instances it may be useful to scale only some of the inputs or minimize cost for a subset of them. We refer to these cases as *Subvector Efficiency Measures*.

To illustrate the subvector idea, consider expression (22), where both inputs are scaled by l. Now if, for example, x_2 is not adjustable (for example this may be the physical plant which is fixed in the short run), then we may drop the l in front of that input and only scale on x_1. Clearly any of the technical efficiency measures can be generalized to scale on subvectors, which is done by multiplying only the relevant inputs by l.

Again, in the short run, some inputs may be fixed and cost may be computed for the variable factors, i.e., short run variable cost. This generalization of the cost minimization problem (38) is obtained by minimizing cost for some inputs, while keeping the others fixed. In our example (37), one may take x_2 equal to the observed input and minimize variable cost with respect to the first input.

3. Direct Output-Oriented Efficiency Measures

Output-oriented measures of efficiency tell you how much more can be produced from a given amount of inputs or resources. In contrast to the input-saving measures where the idea is to reduce inputs; inputs are taken as given here, and outputs are expanded. Nonetheless, there are a lot of similarities: like the input measures, we can divide output measures of efficiency into technical and 'economic' types, where the latter require information on prices, this time output prices. This section will in fact mirror our section on input measures of efficiency. Again, we begin with the technical measures of efficiency.

We begin with a diagram and some numbers to illustrate output-oriented technical efficiency. Since we allow for more than one output, the most helpful way to look at technology is with an output set. In Figure 7 we reproduce the output set from Figure 3 and augment it with the data in Table 6 below. With the additional data there are three observations or DMUs, each using one unit of input to produce two different outputs, y_1 and y_2. DUM 1 and 2 are benchmark firms; they are on the boundary or best practice frontier of the technology,

$P(1 \mid C,S)$. DMU 3, which has the same inputs as 1 and 2 but produces less of both outputs, is in the interior of the output set, and is obviously not as productive as 1 and 2.

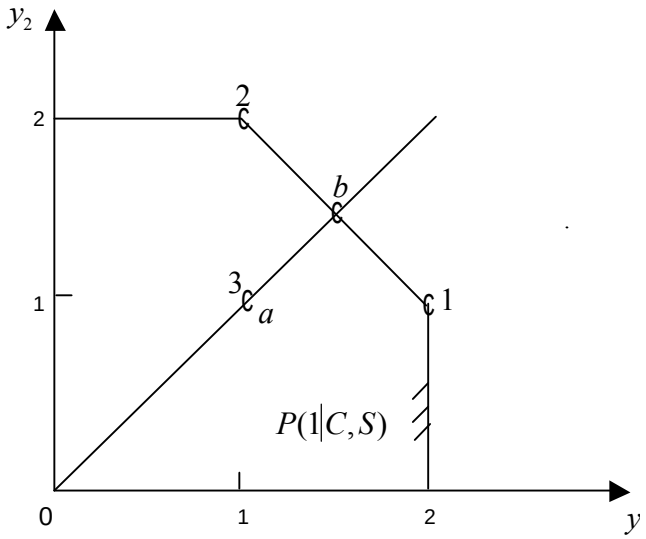

Figure 7: The Output Measure of Technical Efficiency

If we measure the deviation of DMU 3 from the best practice frontier in a radial way, its relative technical efficiency is given by

$$0b / 0a,$$

which can also be thought of as the ratio of (the size of) maximum potential output (at *b*) to (the size of) actual or observed output (at *a*). This measure is sometimes referred to as the *Farrell Output-Oriented Measure of Technical Efficiency*. More formally we define it as

$$F_o(x, y \mid C,S) = \max\{q : qy \in P(x \mid C,S)\}. \tag{42}$$

In order to distinguish between the output and input oriented measures, we use the subscript 'o' instead of 'i'. In fact, the two measures are, as you might guess, related. Specifically, under constant returns to scale and strong disposability, they are reciprocals, i.e.,

$$F_o(x, y \mid C,S) = (F_i(y, x \mid C,S))^{-1}. \tag{43}$$

Firm (DMU)	Input x	Output 1 y_1	Output 2 y_2
1	1	2	1
2	1	1	2
3	1	1	1

Table 6: Augmented Data for an Output Set

We can compute the output-oriented measure of technical efficiency for DMU 3 using our data as follows

$$\begin{aligned} F_o(1,1,1 \mid C,S) = \max & q \\ \text{s.t.} \quad & z_1 2 + z_2 1 + z_3 1 \geqq q 1, \\ & z_1 1 + z_2 2 + z_3 1 \geqq q 1, \\ & z_1 1 + z_2 1 + z_3 1 \leqq 1, \\ & z_1 \geqq 0, z_2 \geqq 0, z_3 \geqq 0, \end{aligned} \tag{44}$$

and the efficiency score is equal to

$$F_o(1,1,1 \mid C,S) = 1.5,$$

which can be interpreted as saying that firm 3 could increase its two outputs by 50% if it were operating on the best practice frontier.

If we computed the efficiency score for DMU 1 then the result is

$$F_o(1,2,1 \mid C,S) = 1,$$

i.e., firm 1 is technically efficient – it is a benchmark firm. Inefficient firms have output efficiency scores greater than one and efficient firms have scores equal to one. In general then

$$F_o(x,y \mid C,S) \geqq 1, \tag{45}$$

and we say that a firm or DMU is technically output efficient if

$$F_o(x,y \mid C,S) = 1, \tag{46}$$

and inefficient otherwise.

Output Slack

Since the output measure of technical efficiency is radial, it is possible that efficient as well as inefficient firms may have what is called *Output Slack*. To see this, suppose we add one more firm to our data, see Table 7.

Firm (DMU	Input x	Output 1 y_1	Output 2 y_2
4	1	0.5	2

Table 7: Added Data

As you can see in Figure 8, DMU 4 is situated on a 'flat spot' of the output set. It uses the same inputs as DMU 2 and produces the same amount of output y_2, but less of y_1. However, if we compute the Farrell measure of output technical efficiency for DMU 4, we would find that its value is 1, i.e., it is technically efficient since we cannot radially expand outputs. Nonetheless, compared to DMU 2, we can see that its production is in some sense smaller, in fact it is smaller by 0.5 units of output one. In general we say that there is *Slack* in output y_m for firm k' if

$$\sum_{k=1}^{K} z_k y_{km} > y_{k'm} \cdot F_o(x^{k'}, y^{k'} \mid C, S), \tag{47}$$

is true for some solution value for $z_k, k = 1,...,K$. In our example, there is slack in output y_1.

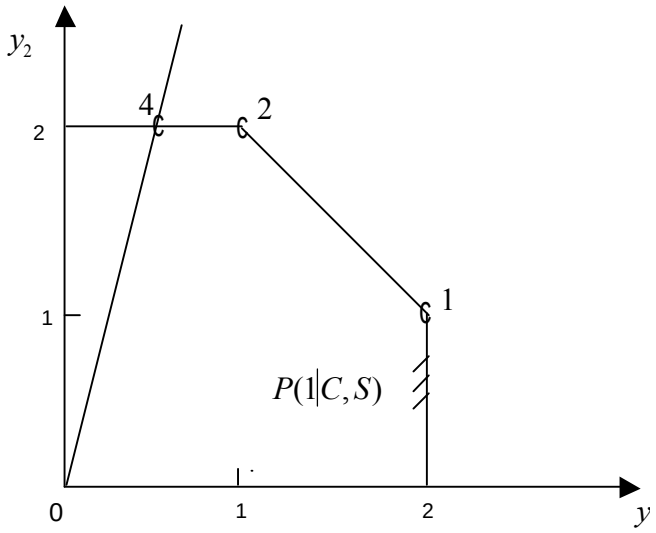

Figure 8: Output Slack

Decomposing Output Technical Efficiency

Next we show how to decompose the output measure of technical efficiency $F_o(x, y \mid C, S)$ into a scale efficiency measure and a measure of congestion. We begin with scale efficiency.

Output Scale Efficiency

We proceed by noting that you can impose various types of returns to scale on the reference technology by changing the restrictions on the intensity variables $(z_1,...,z_K)$. For example if

$$\sum_{k=1}^{K} z_k \leqq 1,$$

then the technology described by

$$\begin{aligned} P(x \mid N, S) = \{(y_1, \ldots, y_M) : & \\ & \sum_{k=1}^{K} y_{km} \geqq y_m, m = 1, \ldots, M, \\ & \sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1, \ldots, N, \\ & \sum_{k=1}^{K} z_k \leqq 1, z_k \geqq 0, k = 1, \ldots, K\}, \end{aligned} \tag{48}$$

satisfies *Nonincreasing Returns to Scale* (N). If, instead the sum of the intensity variables is restricted to exactly equal one, we can model *Variables Returns to Scale* (V),

$$\begin{aligned} P(x \mid V, S) = \{(y_1, \ldots, y_M) : & \\ & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, m = 1, \ldots, M, \\ & \sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1, \ldots, N, \\ & \sum_{k=1}^{K} z_k = 1, z_k \geqq 0, k = 1, \ldots, K\}, \end{aligned} \tag{49}$$

which allows for increasing, constant and decreasing returns to scale.

We can measure output-oriented technical efficiency relative to any of these technologies, for example, we can define

$$F_o(x, y \mid V, S) = \max\{\theta : \theta y \in P(x \mid V, S)\}, \tag{50}$$

which measures technical efficiency relative to a variable returns to scale technology. The ratio of this particular measure and $F_o(x, y \mid C, S)$ is used to define *Output Scale Efficiency*

$$S_o(y, x \mid S) = F_o(x, y \mid C, S) / F_o(y, x \mid V, S), \tag{51}$$

which is a measure of the deviation from constant returns to scale in the output direction. This is easiest to see in a diagram, see Figure 9.

In the figure, firm 1 produces one unit of output from one unit of input. Firm 2 produces 1.5 units of output from 2 units of input. If production were subject to constant returns to scale, firm 2 should be able to produce 2 units of output from 2 units of input, i.e., double the inputs and outputs of firm 1. Note that both firm 1 and 2 are technically efficient relative to the variable

returns to scale technology (labeled VRS in Figure 9), but only firm 1 is efficient when the reference technology satisfies constant returns to scale (labeled CRS). So firm 1 is what we call scale efficient. Firm 2 is not scale efficient; if constant returns to scale prevailed, firm 2 should be able to scale up its output by .5 units, or a proportion of 4/3.

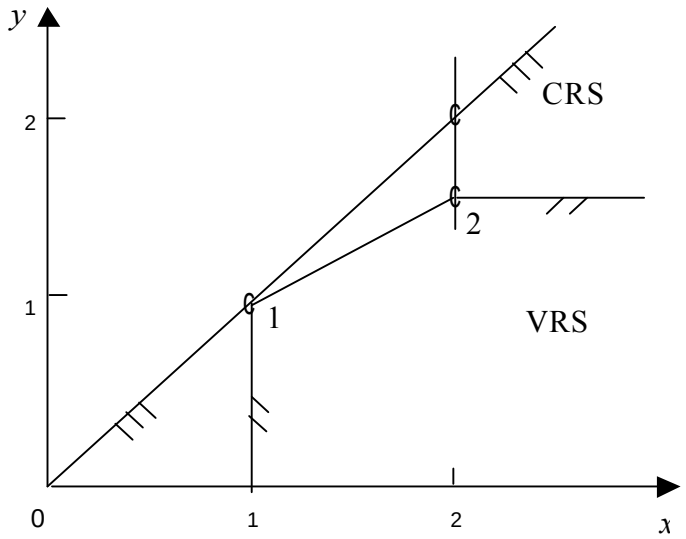

Figure 9: Output Scale Efficiency

More generally, we say that a firm or DMU is scale efficient if $S_o(x,y \mid S)=1$ i.e., if $F_o(x,y \mid C,S)=F_o(x,y \mid V,S)$, otherwise it is scale inefficient, $S_o(x,y \mid S)>1$. Deviations from scale efficiency are essentially deviations from constant returns and therefore can be due to operating at a point of increasing returns or a point of decreasing returns to scale. To identify which is the case, we need to compute technical efficiency relative to a nonincreasing returns to scale technology, i.e.,

$$F_o(x,y \mid N,S)=\max\{\theta : \theta y \in P(x \mid N,S)\}. \tag{52}$$

As it turns out (and is easily seen in Figure 9 if you notice that the boundary of the nonincreasing returns technology overlaps the VRS technology, but includes the ray from point 1 to the origin), we have the following relationship

$$1 \leqq F_o(x,y \mid V,S) \leqq F_o(x,y \mid N,S) \leqq F_o(x,y \mid C,S), \tag{53}$$

and we can identify deviations from scale efficiency according to the following rule

$$\begin{aligned} &\text{if } S_o(x,y \mid S)>1, \quad \text{then scale inefficiency is due to:} \\ &\qquad \text{increasing returns to scale if } F_o(x,y \mid N,S)=F_o(x,y \mid C,S), \\ &\qquad \text{decreasing returns to scale if } F_o(x,y \mid N,S)>F_o(x,y \mid C,S). \end{aligned} \tag{54}$$

In Figure 9, it is easy to see that DMU is scale inefficient and that it is due to operating at a point of decreasing returns. For firm 2 we have

$$F_o(2,1.5 \mid V,S) = 1 = F_o(2,1.5 \mid N,S) < 4/3 = F_o(2,1.5 \mid C,S), \tag{55}$$

so firm 2 is scale inefficient with

$$S_o(2,1.5 \mid S) = 4/3.$$

which is due to decreasing returns to scale since

$$F_o(2,1.5 \mid C,S) > F_o(2,1.5 \mid N,S).$$

We can use scale efficiency to decompose our original output-oriented measure of technical efficiency

$$F_o(x,y \mid C,S) = S_o(x,y \mid S) \cdot F_o(x,y \mid V,S). \tag{56}$$

This decomposition shows that a scale efficiency measure can be extracted from $F(y,x \mid C,S)$; next we show how to also identify a congestion component.

Output Congestion

We use output congestion to mean deviations from strong disposability of outputs. This is often associated with the idea of joint production of 'good' outputs with undesirable byproducts. Typically these undesirable outputs are thought of as 'waste', however, getting rid of these 'bads' is generally costly; either directly through fines or regulations, or indirectly, because their reduction implies joint reduction of good outputs.

In order to model this idea that disposal of jointly produced bads is not free (not strongly disposable),we instead use the idea of weak disposability of outputs. The simplest case would be that in which all outputs are weakly disposable, then we would measure efficiency relative to a technology satisfying weak disposability of all outputs.

$$F_o(x,y \mid V,W) = \max\{\theta : \theta y \in P(x \mid V,W)\}, \tag{57}$$

where the reference technology can be specified as

$$P(x \mid V,W) = \{(y_1,...,y_M) : \sum_{k=1}^{K} z_k y_{km} = s y_m, m = 1,...,M,$$

$$\sum_{k=1}^{K} z_k x_{kn} \leqq x_n, n = 1,..., N,$$
$$\sum_{k=1}^{K} z_k = 1, z_k \geqq 0, k = 1,..., K,$$
$$0 \leqq s \leqq 1\}.$$

Note that when weak disposability (W) is imposed together with variable returns to scale (V), a scaling factor, s, is introduced to ensure that weak disposability is imposed.

To explicitly account for deviations from strong disposability we can now define an *Output Congestion Measure* for $m = 1,..., M$

$$CN_o(y, x \mid V) = F_o(x, y \mid V, S) / F_o(x, y \mid V, W), \tag{59}$$

and we say that an observation *k* is congestion free if $CN_o(x, y \mid V) = 1$, and that it is congested if the measure is greater than one. Instead of restricting all outputs to be weakly disposable, one could also specify that only a subvector of outputs be restricted to be weakly disposable in $F_o(x, y \mid V, W)$; presumable those that are undesirable or costly to remove.

Combining (51) and (59) we can now summarize the following decomposition of our original output measure of technical efficiency from (42), i.e.,

$$F_o(x, y \mid C, S) = S_o(x, y \mid S) \cdot CN_o(x, y \mid V) \cdot F_o(x, y \mid V, W), \tag{60}$$

The first component measures deviations from constant returns to scale and the second component measures deviations from strong disposability of outputs. The last component is a technical efficiency measure computed relative to a reference technology satisfying variable returns to scale and weak disposability of outputs.

Economic Output Efficiency Measures

This section looks at economic or price-related measures of output efficiency. In order to compute such measures, you must have information on output prices as well as output and input quantities. That is, we assume here that output prices $p^k = (p_{k1},..., p_{km})$ are known for each DMU, $k = 1,..., K$. Using the price and quantity data, we can compute *Total Revenue* as

$$R^k = \sum_{m=1}^{M} p_{km} y_{km} = p^k y^k. \tag{61}$$

This observed revenue, as we shall see, may not necessarily be the same as maximum revenue.

A numerical example might be helpful. Table 8 contains some hypothetical data for three DMUs. Let's assume that they all face the same output prices, $p^k = (1,2)$, $k = 1,2,3$.

Firm (DMU)	Input x	Output 1 y_1	Output 2 y_2	Total Revenue
1	1	2	1	4
2	1	1	2	5
3	1	1	1	3

Table 8: Data for a Best Practice Technology

One obvious economic measure of performance is a comparison of the observed total revenues with *Maximum Revenues*. These are defined as

$$R(x, p \mid C, S) = \max\left\{\sum_{m=1}^{M} p_m y_m : (y_1,...,y_M) \in P(x \mid C, S)\right\}, \tag{62}$$

where the idea is to solve for the output quantities that maximize revenues given output prices and the reference technology. Here we have used the technology satisfying strong disposability and constant returns.

Substituting our data into (62), we have for DMU 3,

$$\begin{aligned} R(1,1,1 \mid C, S) = \max\ & 1y_1 + 2y_2 \\ \text{s.t.}\quad & z_1 2 + z_2 1 + z_3 1 \geqq y_1, \\ & z_1 1 + z_2 2 + z_3 1 \geqq y_2, \\ & z_1 1 + z_2 1 + z_3 1 \leqq 1, \\ & z_1 \geqq 0, z_2 \geqq 0, z_3 \geqq 0. \end{aligned} \tag{63}$$

The solution to this problem is \$5, whereas the observed total revenue for DMU 3 was only \$3, i.e., DMU 3 is in some sense inefficient, which is also illustrated in Figure 10, where we have plotted our data along with the reference technology and observed and maximal revenue. In general we define the *Overall Measure of Output Efficiency* as

$$O_o(x, y, p \mid C, S) = R(x, p \mid C, S) / py, \tag{64}$$

i.e., the ratio of maximum to observed total revenue. This ratio is also sometimes referred to as the *Revenue Measure of Output Efficiency*. For firm or DMU 3, observed revenue was \$3, so we have

$$O_o(1,1,1,1,2 \mid C, S) = 0c / 0a = 5/3, \tag{65}$$

where the letters refer to the points in Figure 10.

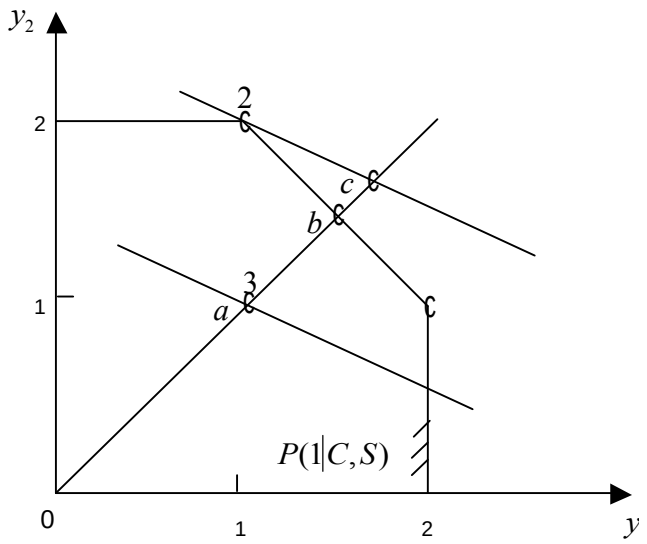

Figure 10: The Output-Oriented Farrell Decomposition

This overall measure can be decomposed along the same lines as the Farrell decomposition on the input side into a technical and allocative part. If you refer to Figure 10, you can see that DMU 3 is not only not earning maximum potential revenue, it is also operating inside the best practice frontier. We can identify the output technical efficiency for firm 3 as

$$F_o(1,1,1 \mid C,S) = 0b / 0a = 1.5 \tag{66}$$

i.e., firm 3 could increase its two outputs by a factor of 1.5 if it were operating at benchmark level *b* instead of at observed level *a*. This component, is of course, independent of prices, although it can be given a revenue interpretation. Specifically, revenue could be increased by a factor of 1.5 if DMU 3 were technically efficient.

Given overall efficiency and technical efficiency, we can now define *Output Allocative Efficiency* as the residual, namely

$$A_o(x,y,p \mid C,S) = O_o(x,y,p \mid C,S) / F_o(x,y \mid C,S)). \tag{67}$$

Rearranging this expression gives the *Farrell Decomposition of Output Efficiency*

$$O_o(x,y,p \mid C,S) = A_o(x,y,p \mid C,S) \cdot F_o(x,y \mid C,S), \tag{68}$$

which is illustrated for DMU 3 in Figure 10, with the associated values

$$0c / 0a = 0c / 0b \cdot 0b / 0a.$$

Or in numbers, for DMU 3 we have

$$5/3 = ((5/3)/(3/2)) \cdot (3/2).$$

If we combine the Farrell decomposition with out decomposition of technical efficiency in (60) we have the following grand decomposition of revenue efficiency

$$O_o(x, y, p \mid C, S) = A_o(x, y, p \mid C, S) \cdot S_o(x, y \mid S) \cdot C_o(x, y \mid V) \cdot F_o(x, y \mid V, W). \qquad (69)$$

Subvector Efficiency

The above output-based technical efficiency measures scale each output with the same factor q. Sometimes it may be useful not to treat output symmetrically and scale only some of them. In this case we have what we call *Subvector Efficiency Measures*. As an example, see expression (44), both outputs are scaled, but if we wish to measure subvector efficiency we just drop the q for the output we do not wish to scale. This simple example generalizes to each of the technical measures: just drop the scaling factor for those outputs you wish to exclude from the scaling.

For revenue maximization, one may maximize over a subvector of outputs. This requires specifying which outputs are to be included in the objective function, and including the observed outputs of those omitted in the objective function on the right hand side of the appropriate output constraint.

4. Measuring Productivity

In this section we measure how performance changes over time. The basic notion we use here is what is typically called productivity or productivity growth. As it turns out, the technical efficiency measures discussed in the two previous sections lend themselves very readily to productivity measurement. In fact they are the natural building blocks for measuring total factor productivity.

To get the basic idea of what we mean by productivity, let's start with simplest possible case: a world in which there is a single output produced by a single input and we have two periods, t and $t+1$. So we observe (x^t, y^t) in the base period and (x^{t+1}, y^{t+1}) in the following period. Total factor productivity is an index of how much output is produced from input(s), so to measure that idea over time we have

$$TFP = \frac{y^{t+1}/x^{t+1}}{y^t/x^t} \qquad (70)$$

which is just the ratio of average products in each period in this simple case. The trick is to construct this type of index when you have more than one input and output, which of course, is the usual case.

In order to generalize to the many input, many output case we follow the index number literature and make use of distance functions to aggregate inputs and outputs. For us the nice thing about distance functions is that they are the reciprocals of the technical efficiency measures discussed earlier in this chapter. We can define the *Input Distance Function* for a $L(y \mid C,S)$ technology as

$$D_i(y,x \mid C,S) = 1/(F_i(y,x \mid C,S)). \tag{71}$$

Similarly, we can define the *Output Distance Function* for technology $P(x \mid C,S)$ as

$$D_o(x,y \mid C,S) = 1/(F_o(x,y \mid C,S)). \tag{72}$$

One of the reasons distance functions have been used in constructing total factor productivity indexes is that they have some nice mathematical properties. These are summarized below.

- By definition, the input and output distance functions are homogeneous of degree plus one in x and y, respectively,

$$\begin{aligned} D_i(y, \lambda x \mid C,S) &= \lambda D_i(y,x \mid C,S), \lambda > 0. \\ D_o(x,\theta y \mid C,S) &= \theta D_o(x,y \mid C,S), \theta > 0. \end{aligned} \tag{73}$$

- In the case where they satisfy constant returns to scale, the input and output distance functions are homogeneous of degree minus one in y and x, respectively,

$$\begin{aligned} D_i(\lambda y,x \mid C,S) &= (1/\lambda) D_i(y,x \mid C,S), \lambda > 0. \\ D_o(\theta x,y \mid C,S) &= (1/\theta) D_o(x,y \mid C,S), \theta > 0. \end{aligned} \tag{74}$$

- When they satisfy constant returns to scale, the input and output distance functions are reciprocals,

$$D_i(y,x \mid C,S) = 1/(D_o(x,y \mid C,S). \tag{75}$$

If we go back to the single input, single output case, we can write the output distance function as follows

$$D_o(x,y \mid C,S) = \frac{y}{x} D_o(1,1 \mid C,S), \tag{76}$$

where we use the properties in (73) and (74). If we use this idea and define something like (76) for t and $t+1$, and substitute into our definition of total factor productivity we get

$$TFP = \frac{y^{t+1}/x^{t+1}}{y^t/x^t} = D_o^t(x^{t+1}, y^{t+1} \mid C,S) / D_o^t(x^t, y^t \mid C,S), \tag{77}$$

where $D_o^t(\cdot)$ is the distance function defined relative to the reference technology from period t.

This type of total factor productivity index can be defined for the general many input and many output case as well, and is called *the Period t Output-Oriented Malmquist Productivity Index*:

$$M_o^t = D_o^t(x^{t+1}, y^{t+1} \mid C,S) / D_o^t(x^t, y^t \mid C,S). \tag{78}$$

This index compares data from two different periods, t and $t+1$, to the same reference technology from period t (note the superscripts on the D as well as x, y in the definition). Figure 11 shows what is going on for a simple example. The observed data from period t is $(x^t, y^t) = (1,1)$ and for period $t+1$ it is $(x^{t+1}, y^{t+1}) = (1.5,2)$. The two distance functions can be computed as

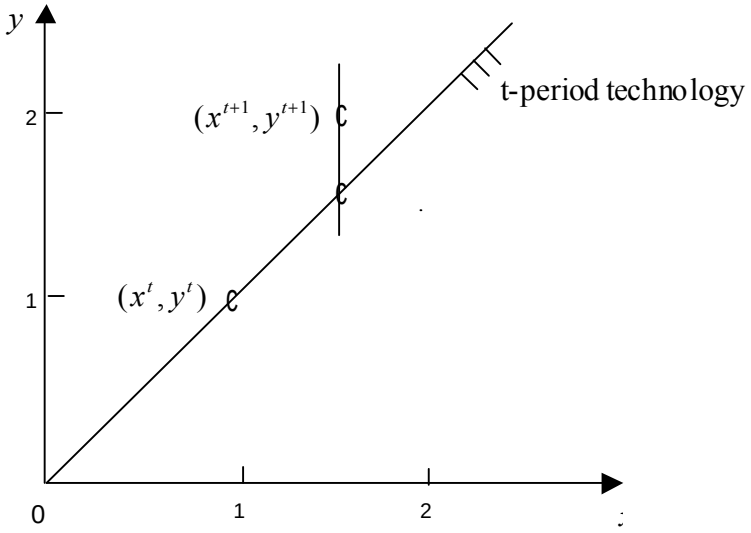

Figure 11: t-Period Malmquist Productivity Index

$$\begin{aligned} (D_o^t(1,1 \mid C,S))^{-1} = \max & \; q \\ \text{s.t.} \quad & z1 \geqq q\,1, \\ & z1 \leqq 1, \\ & z \geqq 0, \end{aligned} \tag{79}$$

and

$$\begin{aligned} (D_o^t(1.5,2 \mid C,S))^{-1} = \max & \; q \\ \text{s.t.} \quad & z1 \geqq q\,2, \end{aligned} \tag{80}$$

$$z1 \leq 1.5,$$
$$z \geq 0,$$

with solution values of $D_o^t(1,1 \mid C,S) = 1$ and $D_o^t(1.5,2 \mid C,S) = 2/(1.5)$. Thus the productivity index is

$$M_o^t = \frac{2}{1.5} = 4/3. \tag{81}$$

Before going on, we note the following about our example:

- The reference technology is constructed from the data from period t only. Here, the data from period $t+1$ lies above the best practice frontier from period t.

- For the computations, we actually compute technical efficiency measures and use the fact that they are reciprocal to the distance functions. This allows us to compute a simple linear programming problem.

- The value of M_o^t is greater than one. In this case that means that there has been an improvement in productivity between period t and $t+1$.

In the example above we define productivity relative to the period t best practice frontier. We can define an analogous productivity measure where the best practice frontier from period $t+1$ is used as the benchmark, namely, *the Period $t+1$ Malmquist Productivity Index* is defined as

$$M_o^{t+1} = D_o^{t+1}(x^{t+1}, y^{t+1} \mid C,S) / D_o^{t+1}(x^t, y^t \mid C,S). \tag{82}$$

We can make use of the t and $t+1$ versions of the Malmquist index to form an 'ideal' type index. This type of index is due to Fisher (1922). The Fisher ideal index is the geometric mean of a Paasche index and a Laspeyres index, which are the upper and lower bounds of the 'true' index. Taking the geometric means of these bounds thus gives a closer approximation to the true index. We use that same idea and take the geometric mean of the t and $t+1$ Malmquist indexes to define the *Output-Oriented Malmquist Productivity Index* (M_o) as

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = \left(\frac{D_o^t(x^{t+1}, y^{t+1} \mid C,S}{D_o^t(x^t, y^t \mid C,S)} \frac{D_o^{t+1}(x^{t+1}, y^{t+1} \mid C,S)}{D_o^{t+1}(x^t, y^t \mid C,S)} \right)^{1/2}. \tag{83}$$

This form of the Malmquist index is illustrated in Figure 12. There are two different best practice frontiers in the figure, one formed from period t data and the other from period $t+1$ data. Included in the figure is data from each period for one DMU, denoted by (x^t, y^t) and

(x^{t+1}, y^{t+1}). The observed input and output from $t+1$ lie 'above' the period t best-practice technology.

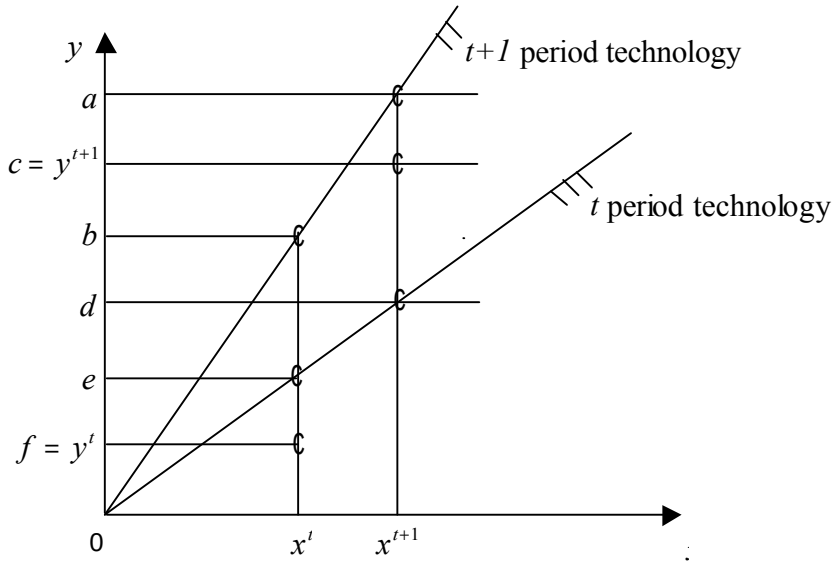

Figure 12: The Output-Oriented Malmquist Productivity Index

If we substitute the letters on the y-axis (recall that this is an output-based measure), the Malmquist index for our DMU in Figure 12 equals

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = \left(\frac{0c/0d}{0f/0e} \frac{oc/oa}{of/ob} \right)^{1/2}. \tag{84}$$

We can rewrite this expression as

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = \left(\frac{0c/0a}{0f/0e} \right) \left(\frac{0a/0d}{0b/0e} \right)^{1/2}, \tag{85}$$

where the expression in the first parentheses measures the change in efficiency between period t and $t+1$: $(0c/0a)$ is the technical efficiency of (x^{t+1}, y^{t+1}) relative to the period $t+1$ best practice frontier and $(0f/0e)$ is the technical efficiency of (x^t, y^t) relative to the t period best practice frontier. We call this term the *Efficiency Change* component of productivity change. In general it is defined as

$$EFFCH = D_o^{t+1}(x^{t+1}, y^{t+1} \mid C, S) / D_o^t(x^t, y^t \mid C, S). \tag{86}$$

The square root of the second term in parentheses in (85) captures the shift in the best practice frontier between t and $t+1$: $(0a/0d)$ measures the vertical shift at x^{t+1} and $(0b/0e)$ captures

the vertical shift evaluated at x^t. The (geometric) mean of these two shifts is our measure of technical change. In general we define *Technical Change* as

$$TECH = \left(\frac{D_o^t(x^{t+1}, y^{t+1} \mid C,S)}{D_o^{t+1}(x^{t+1}, y^{t+1} \mid C,S)} \frac{D_o^t(x^t, y^t \mid C,S)}{D_o^{t+1}(x^t, y^t \mid C,S)} \right)^{1/2}. \tag{87}$$

The produce of *EFFCH* and *TECH* is equal to $M_o(x^{t+1}, y^{t+1}, x^t, y^t)$. Improvements in productivity over time is signaled when the value of $M_o(x^{t+1}, y^{t+1}, x^t, y^t)$ is greater than one, whereas declines in productivity are signaled when its value is less than one. The same interpretation applies to the components of productivity change, *EFFCH* and *TECH*. Note that improvement in productivity could be accompanied by deterioration in one of the component measures, and vice versa.

In the single input, single output case, the Malmquist productivity index simplifies to the total factor productivity measure in (77). In this case, the distance functions simplify to

$$D_o^{t+1}(x^{t+1}, y^{t+1} \mid C,S) = \frac{y^{t+1}}{x^{t+1}} D_o^{t+1}(1,1 \mid C,S) \tag{88}$$

and

$$D_o^t(x^t, y^t \mid C,S) = \frac{y^t}{x^t} D_o^t(1,1 \mid C,S). \tag{89}$$

If we insert these into (83), we get our simple total factor productivity ratios.

There are some potential problems in computing these productivity indexes. Even though we know that the distance functions are the reciprocals of technical efficiency measures, we have some special cases where data from one period is compared to a frontier from a different period, for example $D_o^t(x^{t+1}, y^{t+1} \mid C,S)$ and $D_o^t(x^t, y^t \mid C,S)$. It is possible that these mixed period distance functions may have a solution value of zero, in which case the productivity index will be ill-defined. One way to avoid this problem is to require that all observations are strictly positive, i.e., $x_{kn} > 0$ and $y_{km} > 0$, $k = 1,...,K$, $n = 1,...,N$ and $m = 1,...,M$.

Let's look at the programming problem we have to solve for one of these mixed period problems. We assume that there are $k = 1,...,K$ observations of inputs and outputs at each period t and $t+1$, denoted as $x^{k,t} = (x_{k1}^t,...,x_{kN}^t)$, $x^{k,t+1} = (x_{k1}^{t+1},...,x_{kN}^{t+1})$, $y^{k,t} = (y_{k1}^t,...,y_{kM}^t)$, and $y^{k,t+1} = (y_{k1}^{t+1},...,y_{kM}^{t+1})$. For each observation $k' = 1,...,K$ we compute

$$(D_o^t(x^{k',t+1}, y^{k',t+1} \mid C,S))^{-1} = \max q \tag{90}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{k=1}^{K} z_k y_{km}^{t} \geqq q y_{k'm}^{t+1}, m = 1,..., M, \\ & \sum_{k=1}^{K} z_k x_{kn}^{t} \leqq x_{k'}^{t+1}, n = 1,..., N, \\ & z_k \geqq 0, k = 1,..., K. \end{aligned}$$

The reference technology is formed from the data from period t, which are the summed terms in the problem above. The observation or DMU k' which is under evaluation is, however, data from period $t+1$, as you can see from the superscripts on the right hand side of the inequalities.

The other mixed period distance function is computed for DMU k' as

$$\begin{aligned} (D_o^{t+1}(x^{k',t}, y^{k',t} \mid C,S))^{-1} = & \max q \\ \text{s.t.} \quad & \sum_{k=1}^{K} z_k y_{km}^{t+1} \geqq q y_{k'm}^{t}, m = 1,..., M, \\ & \sum_{k=1}^{K} z_k x_{kn}^{t+1} \leqq x_{k'n}^{t}, n = 1,..., N, \\ & z_k \geqq 0, k = 1,..., K. \end{aligned} \tag{91}$$

Here the reference technology is constructed from the period $t+1$ data, while DMU k' is being evaluated based on its period t data.

Next we look at a numerical example based on the data in Table 9.

Firm (DMU)	Input 1 x^t	Output y^t	Input x^{t+1}	Output y^{t+1}
1	1	1.5	2	3
2	1	1	2	3

Table 9: Two Period Data Set

This data is illustrated in Figure 13.

We will write out the programming problems for the two observations for the mixed period problem $D_o^{t+1}(x^t, y^t \mid C,S)$. For firm 1, we have

$$\begin{aligned} D_o^{t+1}(1,1.5 \mid C,S))^{-1} = & \max q \\ \text{s.t.} \quad & z_1 3 + z_2 3 \geqq q 1.5, \\ & z_1 2 + z_2 2 \leqq 1, \\ & z_1 \geqq 0, z_2 \geqq 0, \end{aligned} \tag{92}$$

which has a solution value of 1. For firm 2

$$(D_o^{t+1}(1,1 \mid C,S))^{-1} = \max q \qquad (93)$$
$$\text{s.t.} \quad z_1 3 + z_2 3 \geqq q 1$$
$$z_1 2 + z_2 2 \leqq 1,$$
$$z_1 \geqq 0, z_2 \geqq 0,$$

which has a solution value of 1.5.

The productivity scores for the two firms are given in Table 10. Firm 2 was technically inefficient in period t, but efficient in period $t+1$, which is reflected in the improvement in *EFFCH*, which is in that case the source of productivity improvements for firm 2. Firm 2 was a benchmark firm in both periods, and, although output increased in period $t+1$, average product remained the same, i.e., there was no technical change, *TECH* = 1.

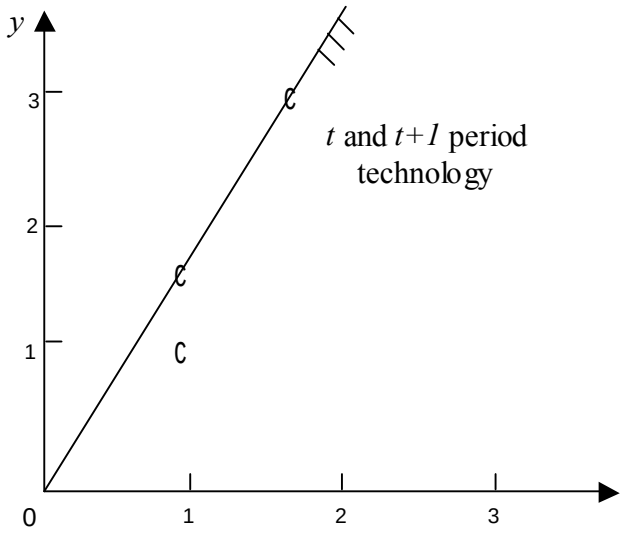

Figure 13: Malmquist Productivity With Sample Data

Firm	$M_o(x^{t+1}, y^{t+1}, x^t, y^t)$	EFFCH	TECH
1	1.0	1.0	1.0
2	1.5	1.5	1.0

Table 10: Productivity Scores

Input-Saving Malmquist Productivity Indexes

Under constant returns to scale, the input and output distance functions are reciprocals. This means that if you would like to measure productivity as input-saving rather than output-

enhancing, all you need to do is invert the output-oriented measure discussed above. That is if we define the *Input-Saving Malmquist Productivity Measure* as

$$M_t(x^{t+1}, y^{t+1}, x^t, y^t) = \left(\frac{D_i^{t+1}(y^{t+1}, x^{t+1} \mid C, S)}{D_i^{t+1}(y^t, x^t \mid C, S)} \frac{D_i^t(y^{t+1}, x^{t+1} \mid C, S)}{D_i^t(y^t, x^t \mid C, S)} \right)^{1/2} \tag{94}$$

then we get

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = 1/(M_i(x^{t+1}, y^{t+1}, x^t, y^t)). \tag{95}$$

In terms of an illustration, if you look at Figure 12, instead of measuring distances North-South, i.e., in an output direction, the input-saving Malmquist index would measure everything East-West, i.e., in an input-saving direction. Again, under constant returns to scale, these are merely reciprocals.

5. Capacity Utilization

Our measure of capacity and capacity utilization has its roots in Johansen (1968, p. 50), who defines plant capacity as '… the maximum amount that can be produced per unit of time with the existing plant and equipment *provided that the availability of the variable factors is not restricted.*' To make this concrete, let the input vector be partitioned into two subvectors $x = (x_f, x_v)$, where x_f is the subvector of fixed factors and x_v is the subvector of variable factors. Define the production function for the single output case $(M = 1)$ as

$$f(x_f, x_v) = f(x) = \max\{y : y \in P(x)\},$$

and define capacity as

$$\hat{f}(x_f) = \max\{f(x_f, x_v) : x_v \geq 0\}$$

which gives the maximum feasible output given x_f, when x_v is unrestricted.

Then the Johansen capacity utilization ratio is given by

$$CU(y, x) = f(x_f, x_v) / \hat{f}(x_f) = f(x) / \hat{f}(x_f), \tag{96}$$

which will be less than or equal to one.

Due to the simple relationship between the output distance function and the production function, namely

$$D_o(x,y) = y / f(x),$$

capacity utilization (96) can be written in terms of output distance functions as

$$CU(y,x) = \hat{D}_o(x_f, y) / D_o(x,y), \tag{97}$$

where

$$\hat{D}_o(x_f, y) = \min\{q : \frac{y}{q} \in P(x_f, x_v), x_v \geqq 0\}.$$

We take (97) as the multi-output measure of the Johansen capacity utilization measure, since the distance functions allow us to easily include a vector of outputs.

The computation of $D_o(x,y)$ is straightforward – it is the reciprocal of the Farrell output oriented technical efficiency measure, $F_o(x,y)$, see for example (44) or (90).

The computation of $\hat{D}_o(x_f, y)$ is also fairly straightforward, namely for observation k',

$$\left(\hat{D}_o(x_f^{k'}, y^{k'} \mid C,S)\right)^{-1} = \max q \tag{98}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} z_k y_{km} \geqq q y_{k'm}, m = 1,...,M$$

$$\sum_{k=1}^{K} z_k x_{k,f} \leqq x_{k'f}, f = 1,...,F$$

$$z_k \geqq 0, k = 1,...,K.$$

Note that in this problem, the fixed factors are restricted, but the variable inputs x_v are not (there is no constraint for x_v).

The solution to (98) may be used to compute capacity output y^* for k' as $y^* = y^{k'} / \hat{D}_o(x_f^{k'}, y^{k'} \mid C,S)$, where $y^{k'}$ may be a vector. One may also compute optimal variable input usage by multiplying the solution values of the intensity (z) variables by the vector of observed variable inputs from the sample, i.e.,

$$x_v^* = \sum_{k=1}^{K} z_k^* x_{kv}, v = 1,...,V.$$

References

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

F@re, R., and S. Grosskopf (1996) *Intertemporal Production Frontiers*, Boston: Kluwer Academic Publishers.

Johansen, L. (1968) Production Functions and the Concept of Capacity, Recherches Recentes sur le Fonction de Production, Collection, *Economie*, Mathematique et Econometric, Vol. 2.

Subject Index

Screen Notation

$F_i(y,x)$	$F_o(x,y)$	Returns to Scale	Disposability
$S_i(y,x)$	$S_o(x,y)$	CRS=C	Strong = S
$CN_i(y,x)$	$CN_o(x,y)$	NIRS=N	Weak = W
$M_i(y,x)$	$M_o(x,y)$	VRS=V	
$A_i(y,x,w)$	$A_o(x,y,p)$	Capacity: Input Designation	
$O_i(y,x,w)$	$O_o(x,y,p)$	fixed= f	
$C(y,w)$	$R(x,p)$	variable= v	
$CU(y,x)$			
$Dhat(x_f,y)$			

EMQ AB
Box 2134, SE-22002 Lund, Sweden
Fax +46 (0)40 164511 Email emq@emq.com Internet www.emq.com