Chapter 6
Productivity Interpretations of the Farrell Efficiency Measures and the Malmquist Index and Its Decomposition

Finn R. Førsund

Abstract The ratio definition of efficiency has the form of a productivity measure. But the weights are endogenous variables and they do not function well as weights in a productivity index proper. It is shown that extended Farrell measures of efficiency can all be given an interpretation as productivity measures as observed productivity relative to productivity at the various projection points on the frontier. The Malmquist productivity index is the efficiency score for a unit in a period relative to the efficiency score in a previous period, thus based on a maximal common expansion factor for outputs or common contraction factor for inputs not involving any individual weighting of outputs or inputs, as is the case if a Törnqvist or ideal Fisher index is used. The multiplicative decomposition of the Malmquist productivity index into an efficiency part and a frontier shift part should not be taken to imply causality. The role of cone benchmark envelopments both for calculating Malmquist indices of productivity change and for decomposing the indices into an efficiency change term and a frontier shift term is underlined, and connected to the index property of proportionality and circularity, adding the use of a fixed benchmark envelopment. The extended decomposition of the efficiency component by making use of scale efficiency is criticised.

Keywords Farrell efficiency measures · Technically optimal scale · Malmquist productivity index · Decomposition of the Malmquist productivity index

JEL Classification C18 · C43 · C61 · D24

F.R. Førsund (✉)
Department of Economics, University of Oslo, Oslo, Norway
e-mail: finn.forsund@econ.uio.no

© Springer International Publishing AG 2016
J. Aparicio et al. (eds.), *Advances in Efficiency and Productivity*,
International Series in Operations Research & Management Science 249,
DOI 10.1007/978-3-319-48461-7_6

6.1 Introduction

Measuring productive efficiency has been developing the last decades to become an important research strand within the fields of economics, management science and operations research. Two seminal contributions are Farrell (1957) and Charnes et al. (1978). Although the latter paper adopts the efficiency definition of the former the approaches for calculation the measure differ in the two papers. Farrell started out defining a frontier production function as the relevant comparison for measuring productive efficiency for observations of production units and introduced radial measures for the case of constant returns to scale. Charnes et al. (1978), formulating the optimisation problem for estimating the efficiency measure, set up a ratio of weighted outputs on weighted inputs. This approach brought the concept of productivity into the efficiency story. However, although the ratio formally looks like a productivity measure it is not set up to represent a productivity index proper, but to estimate efficiency, i.e. to compare the "productivity" of an observation with the productivity of a benchmark on the best practice frontier using weights that are endogenous. Using these weights, the weighted sum of inputs or outputs will be restricted to 1 (depending on estimating an output- or input-oriented efficiency score), and one or more weight may be zero contrary to what one would want constructing a productivity index proper.

A purpose of the chapter is to elaborate upon the productivity interpretation for the generalised Farrell efficiency measures covering the case of variable returns to scale. We then have technical efficiency measures, scale efficiency measures and a technical measure of productivity, the last two types of measures building upon the old concept of technically optimal scale in production theory. We will also have a closer look at the Malmquist productivity index because it is defined as the ratio of Farrell technical efficiency measures for a unit for two different time periods. A contribution of the chapter is to introduce some relevant concepts to an audience oriented toward DEA.

The chapter is organised as follows. The Charnes et al. (1978) ratio measure and five Farrell efficiency measures are defined in Sect. 6.2[1] and the productivity interpretations of the latter measures discussed for the case of a single output and input, and then generalised to multiple outputs and inputs. The importance of (local) constant returns to scale for productivity measurement is brought out using the elasticity of scale. In Sect. 6.3 the Malmquist index proposed in Caves et al. (1982) is introduced and some basic properties of the index and their consequences for choice of efficiency measures are discussed. The decomposition of productivity change into efficiency change and frontier shift introduced in Nishimizu and Page (1982) is discussed and compared with the decomposition done in Färe et al. (1992, 1994a, c). Section 6.4 offers some conclusions.

[1] Section 6.2 is based on Førsund (2015), Sect. 4.

6.2 Productivity Interpretations of the Farrell Efficiency Measures

6.2.1 The Ratio Definition of the Efficiency Measure

Charnes et al. (1978) relate the ratio idea for defining an efficiency measure to how efficiency is defined in engineering as "the ratio of the actual amount of heat liberated in a given device to the maximum amount that could be liberated by the fuel" (Charnes et al. 1978, p. 430). The optimisation problem set up for deriving the efficiency measure in the case of constant returns to scale (CRS) for a dataset, from a specific time period, is:

$$\begin{aligned} &\text{Max}\, h_{j_0} = \frac{\sum_{r=1}^{s} u_{rj_0} y_{rj_0}}{\sum_{i=1}^{m} v_{ij_0} x_{ij_0}} \quad \text{subject to} \\ &\frac{\sum_{r=1}^{s} u_{rj_0} y_{rj}}{\sum_{i=1}^{m} v_{ij_0} x_{ij}} \leq 1\,,\, j = 1\ldots, j_0, \ldots n,\;\; u_{rj_0},\, v_{ij_0} \geq 0\;\forall r, i \end{aligned} \tag{6.1}$$

Here h_{j0} is the efficiency measure for unit j_0, y_{j0} and x_{j0} are the output and input vectors, respectively, with s outputs and m inputs, number of units is n, and u_{rj0}, v_{ij0} are the weights for unit j_0 associated with outputs and inputs, respectively. These weights are endogenous variables and will be determined in the optimal solution. The constraints on the ratios in the optimisation problem (6.1) require the "productivity" of all units to be equal to or less than 1 using the weights for unit j_0, i.e. the productivity of fully efficient units is normalised to 1. Moreover, the weighted sum of inputs (input orientation) or outputs (output orientation) for the unit j_0 under investigation is normalised to 1 when the fractional programming problem (6.1) is converted to a linear programming problem as shown by Charnes et al. (1978), thus providing a link to the Farrell approach.[2]

6.2.2 The Farrell Suite of Efficiency Measures

Farrell (1957) defined two technical measures of efficiency, the input-oriented measure based on scaling inputs of inefficient units down with a common scalar, projecting the point radially to the frontier keeping observed output constant, and the output-oriented measure scaling outputs of inefficient units up with a common scalar, projecting the point radially to the frontier keeping observed inputs constant. The measures were defined for a frontier function exhibiting constant returns to

[2]Farrell and Fieldhouse (1962) were the first to solve the problem of calculating their efficiency measure by using linear programming.

scale.[3] However, he also discussed variable returns to scale and studied this further in Farrell and Fieldhouse (1962), without explicitly introducing measures reflecting scale properties. This was done in Førsund and Hjalmarsson (1974, 1979), developing a family of five efficiency measures. The latter paper illustrated the measures using a smooth variable returns to scale frontier production function exhibiting an *S*-shaped graph as typical for neoclassical production functions obeying the *Regular Ultra Passum Law*[4] of Frisch (1965).[5] However, the efficiency measures are valid for other types of frontier functions as long as a basic requirement of the variation of the elasticity of scale is fulfilled. In this paper the focus will be on a non-parametric piecewise linear frontier function; the generic DEA model exhibiting variable returns to scale (VRS) having a convex production possibility set, and exhibiting the other properties introduced in Banker et al. (1984).[6]

The family of the five Farrell efficiency measures is illustrated in Fig. 6.1 in the case of the frontier within a non-parametric framework being a piecewise linear convex function (Førsund 1992). The point of departure is the observation $P^0 = (y^0, x^0)$ that is inefficient with respect to the VRS frontier. The reference point on the frontier for the input-oriented measure E_1 with respect to the VRS frontier is $P_1^{VRS} = (y^0, x_1^{VRS})$, and the reference point on the frontier for the output-oriented measure E_2 with respect to the VRS frontier is $P_2^{VRS} = (y_2^{VRS}, x^0)$. A second envelopment is indicated by the ray from the origin being tangent to the point P^{Tops}. (I will return to the interpretation of this point below.) This frontier exhibits constant returns to scale (CRS). The reference points on the frontier are $P_1^{CRS} = (y^0, x_1^{CRS})$ and $P_2^{CRS} = (y_2^{CRS}, x^0)$. The dotted factor ray from the origin to the observation gives the productivity of the observation, and the dotted factor ray from the origin to a reference point on the VRS frontier gives the productivity of this reference point. As is easily seen from Fig. 6.1 the productivity at the CRS envelopment is the maximal productivity obtained on the VRS frontier. Comparing the observation with the reference point $P^{Tops} = (y^T, x^T)$ therefore gives the relative productivity of an observation to the maximal productivity on the VRS frontier. Continuing Farrell's numbering of measures a measure E_3 is introduced covering this measurement and is therefore termed the measure of *technical productivity*.[7]

[3]Farrell (1957) points out that the two measures in the case of constant returns to sale are equal.

[4]The Regular Ultra Passum Law requires that the scale elasticity decreases monotonically from values greater than one, through the value one to lower values when moving along a rising curve in the input space.

[5]This may be the reason for this way of presenting the family of efficiency measures being rather unknown in the DEA literature.

[6]In the VRS DEA specification the scale elasticity has a monotonically decreasing value in the range of increasing returns to scale, but has a more peculiar development in the range of decreasing returns to scale as shown in Førsund et al. (2009). However, there may be a unique face where the scale elasticity is equal to 1 along a rising curve, or else define a vertex point as having constant returns to scale when the left-hand elasticity at the point is less than one and the right-hand elasticity is greater than one.

[7]In Førsund and Hjalmarsson (1979), introducing this measure, it was called the gross scale efficiency.

The two remaining efficiency measures E_4 and E_5 introduced in Førsund and Hjalmarsson (1979) are the scale efficiency measures[8] comparing the productivity of the reference points P_1^{VRS} and P_2^{VRS}, respectively, with the point P^{Tops} of maximal productivity on the frontier.

6.2.3 Productivity Interpretations in the Case of a Single Output and Input

All Farrell measures of efficiency can be given an interpretation of relative productivity; the productivity of the observation relative to specific points on the VRS frontier marked in Fig. 6.1. Before showing the relative productivity interpretation in the case of a single output and a single input (Berg et al. 1992) in a general setting, let us state the definitions of the Farrell input-and output-oriented technical efficiency measures, starting with the general definition of the production possibility set $T = \{(y, x) : y \geq 0 \text{ can be produced by } x \geq 0\}$ (y and x are vectors). By assumption let the set T exhibit variable returns to scale (VRS) of its frontier (the efficient boundary of T). The input-and output-oriented efficiency measures can be defined as[9]

$$\begin{aligned} E_1(y,x) &= \min_{\mu}\{\mu : (\mu x, y) \in T\} \\ E_2(y,x) &= \min_{\lambda}\{\lambda : (x, y/\lambda) \in T\}. \end{aligned} \tag{6.2}$$

The relative productivity interpretation can be shown in the case of a single output and input in the following way, starting with the input-oriented efficiency measure using the points P^0 and P_1^{VRS} in Fig. 6.1:

$$\frac{y^0/x^0}{y^0/x_1^{VRS}} = \frac{y^0/x^0}{y^0/E_1 x^0} = E_1 \tag{6.3}$$

The same productivity interpretation holds for the output-oriented efficiency measure using points P^0 and P_2^{VRS} in Fig. 6.1:

$$\frac{y^0/x^0}{y_2^{VRS}/x^0} = \frac{y^0/x^0}{(y^0/E_2)/x^0} = E_2 \tag{6.4}$$

In the input-oriented case we adjust the observed input quantity so that the projection of the observation is on the frontier, and in the output-oriented case we

[8]In Førsund and Hjalmarsson (1979) these measures were called measures of pure scale efficiency.

[9]The Farrell efficiency measure functions correspond to the concept of distance functions introduced in Shephard (1970). Shephard's input distance function is the inverse of Farrell's input-oriented efficiency measure, and Shephard's output distance function is identical to Farrell's output-oriented efficiency measure.

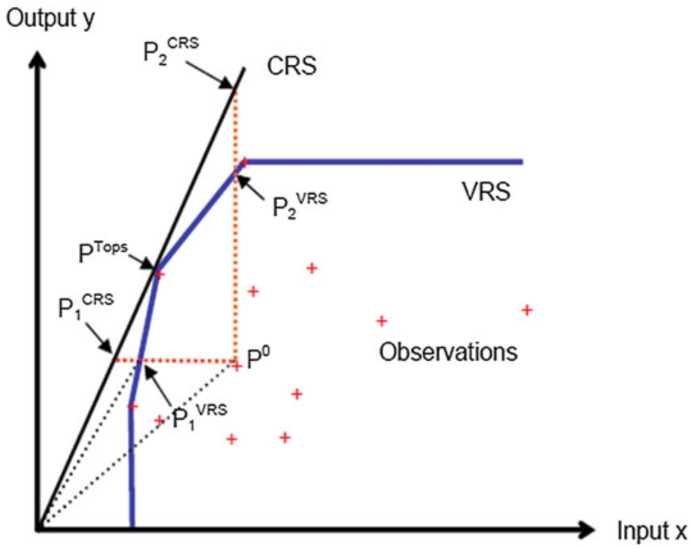

Fig. 6.1 The Farrell efficiency measures applied to a piecewise linear VRS frontier

adjust the observed output, using the symbols for adjusted input and output introduced above.

For the three remaining measures we will make a crucial use of the CRS envelopment in order to calculate the measures. The notation E_1^{CRS} and E_2^{CRS}, making explicit reference to the CRS envelopment as the frontier, together with $\mathrm{P}^{\mathrm{Tops}} = (y^{\mathrm{T}}, x^{\mathrm{T}})$, will be used. The measure of technical productivity is

$$\begin{aligned}
\frac{y^0/x^0}{y^T/x^T} &= \frac{y^0/x^0}{y^0/E_1^{CRS}x^0} = E_1^{CRS} = E_3 \\
\frac{y^0/x^0}{y^T/x^T} &= \frac{y^0/x^0}{(y^0/E_2^{CRS})/x^0} = E_2^{CRS} = E_3 \Rightarrow E_3 = E_1^{CRS} = E_2^{CRS}
\end{aligned} \tag{6.5}$$

The first expression in each of the two lines of the equations is the definition of the measure of technical productivity using the productivity at the point $\mathrm{P}^{\mathrm{Tops}}$ as a reference. The second expressions, input-orientation or output-orientation, respectively, show the most convenient way of calculating the productivity measure. The outputs and inputs differ between the observation P^0 and the $\mathrm{P}^{\mathrm{Tops}}$ points. But using the CRS envelopment the maximal productivity for the VRS technology is the same along the entire ray from the origin going through the point $\mathrm{P}^{\mathrm{Tops}}$. The productivity measure E_3 is equal to both the input-oriented measure and the output-oriented measure using the CRS envelopment as the frontier. It is easy to see geometrically that in the case of using the CRS envelopment the two orientated efficiency measures must be identical, as pointed out by Farrell (1957).

Measures for scale efficiency are also defined using a relative productivity comparison. The input-oriented scale efficiency E_4 (keeping output fixed) and the output-oriented scale efficiency E_5 (keeping input fixed) are:

$$\begin{aligned} \frac{y^0/x_1^{VRS}}{y^T/x^T} &= \frac{y^0/E_1x^0}{y^0/E_1^{CRS}x^0} = \frac{E_1^{CRS}}{E_1} = \frac{E_3}{E_1} = E_4 \\ \frac{y_1^{VRS}/x^0}{y^T/x^T} &= \frac{(y^0/E_2)/x^0}{y^0/E_2^{CRS}x^0} = \frac{E_2^{CRS}}{E_2} = \frac{E_3}{E_2} = E_5 \end{aligned} \tag{6.6}$$

The relative productivity comparison for input-oriented scale efficiency in Fig. 6.1 is between the observed output on the efficiency-corrected input on the VRS frontier and the maximal productivity at the P^{Tops}-point (y^T, x^T). For output-oriented scale efficiency we have an analogous construction. The calculations of the scale efficiency measures can either be based on the ratios between the efficiency scores for input-oriented efficiency relative to the VRS frontier and the CRS envelopment, or expressed as deflating the technical productivity measure with the relevant efficiency measures relative to the VRS frontier. Notice that there is only a single technical efficiency measure for a CRS technology; we have $E_1 = E_2 = E_3$ and $E_4 = E_5 = 1$.

6.2.4 The Concepts of Elasticity of Scale and Technically Optimal Scale

Before generalising the relative productivity interpretation to multiple outputs and inputs we need to introduce the concept of elasticity of scale. The definition of a local scale elasticity for a frontier production function is the same whether it is of the neoclassical differential type $F(y, x) = 0$ or if the production possibility set has a faceted envelopment border like in the DEA case. We are looking at the maximal uniform proportional expansion β of outputs for a given uniform proportional expansion α of inputs, i.e. looking at $F(\beta y, \alpha x) = 0$. The local scale elasticity is defined as the derivative of the output expansion factor w.r.t. the input expansion factor on the average value of the ratio of the output factor on the input factor[10]:

$$\varepsilon(x, y) = \frac{\partial \beta(x, y, \alpha)}{\partial \alpha} \frac{\alpha}{\beta} = \frac{\partial \beta(\alpha, x, y)}{\partial \alpha} \Big|_{\alpha=\beta=1} \tag{6.7}$$

The scale elasticity is evaluated, without loss of generality, for $\alpha = \beta = 1$. In the DEA case with non-differentiable points (vertex points or points on edges) the expression above is substituted with the right-hand derivative or the left-hand derivative, respectively, at such points (Krivonozhko et al. 2004; Førsund and

[10]See Hanoch (1970), Panzar and Willig (1977), Starrett (1977).

Hjalmarsson 2004b; Førsund et al. 2007; Podinovski et al. 2009; Podinovski and Førsund 2010).

Returns to scale is defined by the value of the scale elasticity; increasing returns to scale is defined as $\varepsilon > 1$, constant returns to scale as $\varepsilon = 1$ and decreasing returns to scale as $\varepsilon < 1$.

For a production function with variable returns to scale there is a connection between the input- and output-oriented measures via the scale elasticity. Following Førsund and Hjalmarsson (1979) in the case of a frontier function for a single output and multiple inputs we have

$$E_2 = E_1^{\bar{\varepsilon}} \Rightarrow E_1 \frac{>}{<} E_2 \text{ for } \bar{\varepsilon} \frac{>}{<} 1 \tag{6.8}$$

where the variable $\bar{\varepsilon}$ is the average elasticity of scale along the frontier function from the evaluation point for the input-saving measure to the output-increasing measure. In Førsund (1996) this result was generalised for multiple outputs and inputs in the case of a differentiable transformation relation $F(y,x) = 0$ as the frontier function, using the *Beam [Ray] variation equations* of Frisch (1965). This result holds for points of evaluation being projection points in the relative interior of faces. The path between the points will be continuous although not differentiable at vertex point or points located at edges.

We must distinguish between scale elasticity and scale efficiency (Førsund 1996). Formalising the illustration in Fig. 6.1 the reference for the latter is the concept of *technically optimal scale* of a frontier function (Frisch 1965). The set of points $TOPS^T$ having maximal productivities for the (efficient) border of the production possibility set $T = \{(y,x) : y \geq 0 \text{ can be produced by } x \geq 0\}$ with the frontier exhibiting VRS, can be defined as (Førsund and Hjalmarsson 2004a)[11]

$$TOPS^T = \{(y,x) : \varepsilon(y, x) = 1, (y, x) \in T\} \tag{6.9}$$

It must be assumed that such points exist and that for outward movements in the input space the scale elasticity cannot reach the value of 1 more than once for a smooth neoclassical frontier. However, it can in the DEA case be equal to 1 for points on the same face (see footnote 6). The point (y^T, x^T) used above is now replaced by vectors y^T and x^T belonging to the set $TOPS^T$. From production theory we know that in general a point having maximal productivity must have a scale elasticity of 1. In a long-run competitive equilibrium the production units are assumed to realise the technically optimal scale with the scale elasticity of 1 implying zero profit.

[11]The concept of the M-locus in the case of multi-output was introduced in Baumol et al. (1982, pp. 58–59). In Førsund and Hjalmarsson (2004a) the M locus is defined and estimated within a DEA model using the TOPS set.

6.2.5 The Productivity Interpretation of the Efficiency Measures in the General Case

The interpretation of the five Farrell measures as measures of relative productivity can straightforwardly be generalised to multiple outputs and inputs. Introducing general aggregation functions[12] $Y = g_y\ (y_1,\ y_2,\ldots,y_M)$ and $X = g_x\ (x_1,\ x_2,\ldots,\ x_N)$ where Y and X are the scalars of aggregate quantities and $y_1, y_2,\ldots$ and $x_1, x_2,\ldots$ etc., are elements of the respective vectors y and x for outputs and inputs. The non-negative aggregation functions are increasing in the arguments and linearly homogeneous in outputs and inputs, respectively (O'Donnell 2012). We have, starting with the definition of relative productivity in the input-oriented case for an observation vector $(y^0,\ x^0)$:

$$\underbrace{\frac{g_y(y^0)/g_x(x^0)}{g_y(y_1^{VRS})/g_x(x_1^{VRS})}}_{\text{Relative productivity}} = \underbrace{\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/g_x(E_1x^0)}}_{\text{Substituting for frontier input}} = \underbrace{\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/E_1g_x(x^0)}}_{\text{Using homogeneity of index function}} = E_1 \tag{6.10}$$

In the first expression relative productivity is defined in the input-oriented case using the observed vectors y^0, x^0 and the vectors y_1^{VRS}, x_1^{VRS} for the projection onto the VRS frontier analogous to the point P_1^{VRS} in Fig. 6.1 in the two-dimensional case. In the second expression the vectors for y_1^{VRS} and x_1^{VRS} are inserted, keeping the observed output levels y^0 and contracting the observed input vector using the input-oriented efficiency E_1 to project the inputs x^0 to the VRS frontier. In the third expression the homogeneity property of the input index function is used.

In the case of output orientation of the efficiency measure E_2 we get in the multiple output—multiple input case following the procedure above[13]:

$$\frac{g_y(y^0)/g_x(x^0)}{g_y(y_2^{VRS})/g_x(x_2^{VRS})} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0/E_2)/g_x(x^0)} = \frac{g_y(y^0)/g_x(x^0)}{(g_y(y^0)/E_2)/g_x(x^0)} = E_2 \tag{6.11}$$

Using the general aggregation functions $g_y(y)$, $g_x(x)$ the measure E_3 of technical productivity can be derived using input- or output-orientation:

[12]Following the classical axiomatic (test) approach there are a number of properties (at least 20) an index should fulfil (Diewert 1992), the ones most often mentioned are monotonicity, homogeneity, identity, commensurability and proportionality. "Satisfying these standard axioms limits the class of admissible input (output) quantity aggregator functions to non-negative functions that are non-decreasing and linearly homogeneous in inputs (outputs)" (O'Donnell 2012, p. 257). There is no time index on the functions here because our variables are from the same period.

[13]The productivity interpretation of the oriented efficiency measures E_1 and E_2 can also be found in O'Donnell (2012, p. 259) using distance functions.

$$\frac{g_y(y^0)/g_x(x^0)}{g_y(y^T)/g_x(x^T)} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/g_x(E_1^{CRS}x^0)} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)} = E_1^{CRS} = E_3$$
$$\frac{g_y(y^0)/g_x(x^0)}{g_y(y^T)/g_x(x^T)} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0/E_2^{CRS})/g_x(x^0)} = \frac{g_y(y^0)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)} = E_2^{CRS} = E_3$$
$$\Rightarrow E_1^{CRS} = E_2^{CRS} = E_3 \tag{6.12}$$

We obtain the same relationship between the technical productivity measure and the oriented measures with the CRS envelopment as in the simple case illustrated in Fig. 6.1. Notice that the use of points on the CRS envelopment in (6.12) is just introduced in order to calculate the measure E_3, and is not the basic definition of the measure; the definition is the first expression on the left-hand side of the two first lines.

The case of multi-output and -input is done in the same way for the scale efficiency measures as for the other measures utilising the homogeneity properties of the aggregation functions:

$$\frac{g_y(y^0)/g_x(x_1^{VRS})}{g_y(y^T)/g_x(x^T)} = \frac{g_y(y^0)/g_x(E_1x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)} = \frac{g_y(y^0)/E_1g_x(x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)}$$
$$= \frac{E_1^{CRS}}{E_1} = \frac{E_3}{E_1} = E_4$$
$$\frac{g_y(y_2^{VRS})/g_x(x^0)}{g_y(y^T)/g_x(x^T)} = \frac{g_y(y^0/E_2)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)} = \frac{(g_y(y^0)/E_2)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)}$$
$$= \frac{E_2^{CRS}}{E_2} = \frac{E_3}{E_2} = E_5 \tag{6.13}$$

Again, we obtain the same relationship between the technical productivity measure and the oriented measures defining scale efficiency as in the simple case illustrated in Fig. 6.1. The calculations of the scale efficiency measures can either be based on the ratios between the efficiency scores for input-oriented efficiency relative to the VRS frontier and the CRS envelopment or expressed as deflating the technical productivity measure with the relevant efficiency measures relative to the VRS frontier.

6.3 The Malmquist Productivity Index

The point of departure is that we have observations of a set of the same units over time. The general construction of a total factor productivity index is to have an index for the volume of outputs over a volume index of inputs. A classical problem is to construct appropriate indices aggregating outputs and inputs, respectively.

The special feature of the Malmquist productivity index is that, without having any price data, volume indices can be established based on using efficiency scores for observations relative to estimated frontier production functions representing best practice. Caves et al. (1982) introduced the bilateral Malmquist productivity index developed for discrete time based on the ratio of distance functions (or Farrell efficiency functions that is the term used in this chapter) measured for two observations of the same unit at different time periods utilising efficiency scores only. Färe et al. (1994a, c) showed how to estimate the index in the case of specifying the possibility set as a convex polyhedral set and estimating the border of the set and efficiency scores using linear programming. The popularity soon followed. Caves et al. (1982) have 938 citations and Färe et al. (1994c) 929 in the Web of Social Science (per April 4, 2016).

However, the Malmquist productivity index was oriented, building on either an output-oriented efficiency score or an input-oriented one. A Malmquist index more of the traditional non-oriented type based on an index of output change over an index of input changes for two periods was introduced by Bjurek (1996), inspired by Diewert (1992) mentioning some ideas of Moorsteen and Hicks, hence the name Moorsteen-Hicks index adopted later, although Bjurek used the more functional name of a Malmquist total factor productivity index.[14]

However, a purpose with the present study is to look deeper into the decompositions of the original Caves et al. (1982) Malmquist productivity index, completely dominating in number of applications.[15]

6.3.1 The Interpretation of the Malmquist Productivity Change Index

The Caves et al. Malmquist oriented indices are utilising Farrell technical efficiency scores. The index for a unit i observed for two different time periods u and v, relative to the same border of the production possibility set indexed by b, representing one of the years, is:

$$M_{ij}^{b}(u,v) = \frac{E_j^b(x_{iv}, y_{iv})}{E_j^b(x_{iu}, y_{iu})}, j = 1,2, \quad i = 1,\ldots,N, \quad u,v = 1,\ldots,T, \quad u < v, \quad b = u,v \tag{6.14a}$$

The benchmark technology indexed by b is in many applications either the technology for period u or v, and changing over time according to the technology

[14]A thorough evaluation of the advantages of this type of a Malmquist productivity index is found in Lovell (2003), and it is also mentioned as the most satisfactory Malmquist type of productivity index in O'Donnell (2012), being what he called multiplicatively complete.

[15]Lovell (2003) decomposes also the Malmquist total factor productivity index multiplicatively into five terms. However, we will not investigate this issue here.

chosen as the base for the two periods involved. It is also usual to take a geometric mean of the results using technologies for both year u and v, following the seminal paper Färe et al. (1994a) on how to estimate the Malmquist productivity index.[16] The reason usually given in the literature is simply that either the technology from u or from v may be used as benchmark, and it is arbitrary which one to use, so the most reasonable is to take the geometric mean. As stated in Balk (1998, p. 59): "Since we have no preferences for either the geometric average of these index numbers will be used". Fare et al. (1994c, p.70) stated the reason as "In order to avod choosing an arbitrary benchmark". When a geometric mean is taken technologies for the two periods are involved, and when time moves forward this implies that the technology for a specific period is involved in two productivity change calculations (except for the first and last year).[17]

However, the time periods may be seen to impose a natural choice of the first period as a base in accordance with a "Laspeyres" view of using period u technology to gauge the productivity change from u to v. If the efficiency score for period v is greater (smaller) than the efficiency score for period u using period u technology, then there has been a productivity improvement (deterioration) from period u to period v.

It is well known in the literature how to set up LP problems to estimate the distance (or efficiency) functions involved in (6.14a) so we do not find it necessary to do this here (see e.g. Fried et al. 2008).

The efficiency functions in (6.14a) show the maximal proportional expansion (outputs) or contraction (inputs), and the measures are called technical efficiency measures because prices are not involved. The Malmquist productivity index is then a technical productivity index. There is no aggregation of outputs and inputs involved. Productivity change is measured as the relative change in the common expansion (contraction) factor between two periods.[18]

The productivity results may be different from the results one would get using prices for aggregating outputs and inputs. Weighting with revenue and cost shares as in the Törnqvist index means that the (real) price structure will have an influence. In general it seems more functional to choose weights according to importance placed on variables. The weights appearing in (1) are technically the dual variables, i.e. the shadow prices on output and input constraints (solving the "envelopment problem" in a DEA linear programming model) and give the marginal impact on the efficiency scores of changes in the exogenous observations, and are thus not related to the relative importance in an economic sense. Moreover, these shadow prices changes from one solution to the next in a more or less unpredictable manner. Using

[16]However, no reason is given for this procedure other than claiming that this was done in Caves et al. (1982). But there the geometric mean appears when establishing the connection between the Malmquist index and an Törnquist index assuming the unit to be on the frontier, while the fundamental assumption in Färe et al. (1994a) is that units may be inefficient.

[17]This may explain the empirical result in Bjurek et al. (1998) that productivity developments more or less follow each other for different formulations of the Malmquist index.

[18]The weighted ratio appearing in (1) should not be interpreted as productivity; the weights are just a by-product of the solutions of the optimisation problems in (6.2).

the ratio form as in (6.1) as a productivity index for the development between two time periods means that the weights are different for the solution of the efficiency scores for each period (Førsund 1998).

Another source of difference is that one or more of the weights of outputs and inputs in Eq. (6.1) may be zero, thus excluding variables from explicit influence on the efficiency scores in (6.14a) in order to maximise (minimise) the scaling factors in Eq. (6.2).[19] This may bias the Malmquist index in both directions compared with a standard Törnqvist index where all variables have strictly positive weights.

Another feature of the Malmquist productivity index that may give different results than other indices is that the efficiency functions in (6.14a) are based on frontier functions. In the case of capital vintage effects a dynamic investment process takes place in order to improve the technology level of a firm, so a frontier based on the best vintage technology may give a too optimistic view of the potential for efficiency improvements in the short run (Førsund 2010). The estimation of the frontier using DEA will also be distorted if observations picked to represent best practice by the method may in fact not be best practice, but picked due to biased technical change, as shown in Belu (2015), assuming a single vintage for each unit.

Thus, there is a question about the usefulness of the information a Malmquist productivity index gives compared with indices using available price information. Public sector production activities not selling outputs in markets seem to be the most relevant type of activities for application of the Malmquist productivity index.

In Sect. 6.2 the general aggregator functions $g_y(.)$ and $g_x(.)$ for outputs and inputs was introduced. These functions may now be period-specific. However, because we do not know these or do not have data to estimate them, the Malmquist index will be estimated using non-parametric DEA models giving us the efficiency measures in the numerator and denominator in (6.14a) (Färe et al. 2008).

When applying the Malmquist productivity index attention should be paid to desirable properties. In the literature this is more often than not glossed over. I will therefore explain in more detail the choice of the specification. Productivity as measured by the Malmquist index (6.14a) may be influenced by changes in the scale of the operation, but two units having the same ratio of outputs to inputs should be viewed as equally productive, regardless of the scale of production (Grifell-Tatjé and Lovell 1995). Doubling all inputs and outputs, keeping input and output mixes constant, should not change productivity. Therefore the benchmark envelopment of data, if we want to measure total factor productivity (TFP), is one that is homogenous of degree 1 in the input and output vectors, and thus the linear-homogenous set that fits closest to the data. The homogenous envelopment is based on the concept of technically optimal scale termed TOPS in Sect. 6.2. As pointed out in that section the productivity is maximal at optimal scale where returns to scale is one, thus the CRS contemporary benchmark envelopments

[19]To the best of my knowledge the pattern of occurrence of zero weights in Malmquist productivity index estimations has never been reported in the literature.

(assuming that the contemporaneous frontiers are VRS) are natural references for productivity changes over time.

In Fig. 6.2 observations of the same unit for the two periods *u* and *v* are indicated by P_u and P_v. The two corresponding VRS frontiers are drawn showing an outward shift indicating technological progress. The TOPS point for period *v* is labelled P_v^{Tops}. Just as the productivity should be unchanged if the input and output vectors are proportionally scaled, a measure of productivity should double if outputs are doubled and inputs are kept constant, and increase by half if inputs double, but outputs are constant. The desirable homogeneity properties of a TFP index is therefore to be homogenous of degree 1 in outputs in the second period *v* and of degree (−1) in inputs of the second period, and homogenous of degree (−1) in outputs of the first period *u* and homogenous of degree 1 in inputs of the first period. Using CRS to envelope the data is thus one way of obtaining all the required homogeneity properties of a Malmquist productivity change index. Notice that in the illustration in Fig. 6.2 the relative technology gap between the CRS benchmark technologies (blue lines) for observations in period *v* and *u* are identical, thus making the use of geometric mean of the Malmquist index in (6.14a) superfluous.[20]

The frontier technology level "jumps" from period to period from the start of one period to the start of the consecutive one. Outputs are produced and inputs consumed during the periods. This set-up is of course somewhat artificial compared with the fact that real changes take place in continuous time. The dynamic problems of adapting new technology and phasing it in are neglected. This theme is discussed in e.g. the literature on the Porter hypothesis and environmental regulation (Porter and van der Linde (1995); Brännlund and Lundgren 2009).[21]

Another property of a productivity index held to be important (Samuelson and Swamy 1974) is the *circularity* of the index (Berg et al. 1992; Balk and Althin 1996) (see Gini (1931) for an interesting exposition). The implied transitivity of the index means that the productivity change between two non-adjacent periods can be found by multiplying all the pairwise productivity changes of adjacent periods between the two periods in question, thus making identification of periods with weak or strong productivity growth possible. We will transitivise the Malmquist index by using a single reference frontier enveloping the pooled data, as illustrated by the upper (red) ray CRS(s) in Fig. 6.2. In Tulkens and van den Eeckaut (1995) this type of frontier was termed the *intertemporal frontier*.[22] Notice that taking the

---

[20]Most illustrations of the Malmquist indices in studies using geometric means are in fact using CRS frontiers and single output and input. Considering multiple outputs and inputs distances between contemporaneous frontiers will be independent of where the measure is taken if inverse homotheticity is assumed in addition to CRS, i.e. if Hicks neutral technical change is assumed.

[21]In panel data models efficiency change has been specified (Cornwell et al. 1990) as having unit-specific efficiencies that varies over time, but this is a "mechanical" procedure without an economic explanation of efficiency change.

[22]In Pastor and Lovell (2005), missing out on this reference, it was called the global frontier.

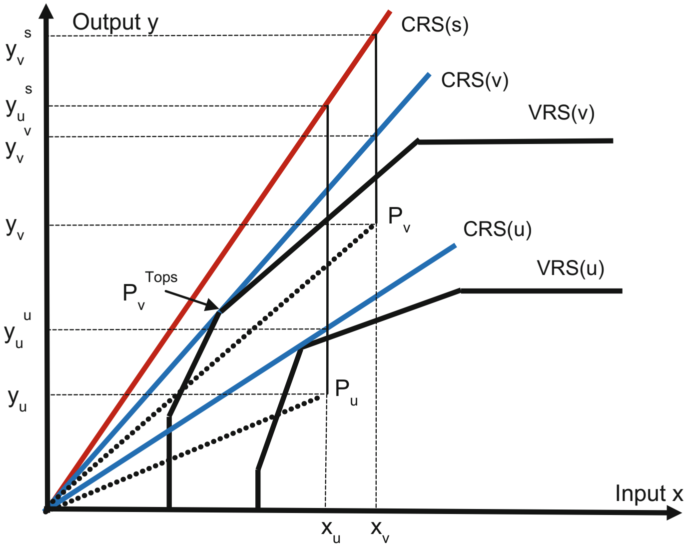

Fig. 6.2 The Malmquist productivity change index. Productivity change for a unit from period *u* to period *v* measured relative to the benchmark CRS(s) envelopment with maximal productivity of the pooled dataset

geometric mean of the Malmquist index (6.14a) for *u* and *v* used as benchmark envelopments is not compatible with circularity.

Using the same CRS reference envelopment for all units means that we have made sure that efficiency for all units and time periods refer to the same envelopment. The observations are either below the benchmark or on it in the case of the units from the pooled dataset spanning the envelopment. The pooled benchmark is identical to the *sequential frontier* of Tulkens and van den Eeckaut (1995) for the last period using sequentially accumulated data of all periods following the argument in Atkinson and Stiglitz (1969) that technology should not be forgotten.

Specifying CRS only is not sufficient to ensure that a specific data point occurring at different time periods get the same efficiency evaluation, because both input- and output isoquants may differ in shape over time if the technology is allowed to change over time as in Färe et al. (2008). The implication of having a time series of data is seldom discussed. Most illustrations and discussions seem to be focussed on two periods only. Hoverer, changing technologies successively as in (6.14a) implies that observations are measured against different frontiers over time. The question is the relevance for estimating productivity change of the information given by comparing relative numbers measured against different benchmarks.

Using a linear homogeneous envelopment implies that the orientation of the E function does not matter. The Malmquist index for a unit *i*, that should be used according to the properties outlined above is then:

$$M_i^s(u,v) = \frac{E^s(x_{iv}, y_{iv})}{E^s(x_{iu}, y_{iu})} = \frac{E_3^s(x_{iv}, y_{iv})}{E_3^s(x_{iu}, y_{iu})}, \quad i = 1, \ldots, J, \quad u, v = 1, \ldots, T, \; u < v \tag{6.14b}$$

where superscript s symbolises that all data are used for estimating the benchmark reference set. The productivity change is the change in the productivities of the observations relative to the benchmark maximal productivity, thus the E^s measures could formally be called E_3^s measures according to the terms introduced in Sect. 6.2, as done in the last expression in (6.14b). If all inputs are increased with a factor α and outputs with factor β from period u to period v ($\alpha x_{iu} = x_{iv}$ and $\beta y_{iu} = y_{iv}$) then we get from (6.14b): $M_i^s(u,v) = E_3^s(\alpha x_{iu}, \beta y_{iu})/E_3^s(x_{iu}, y_{iu}) = \beta/\alpha$; i.e. proportionality is obeyed due to the homogeneity properties of the efficiency score functions.

6.3.2 The Decomposition of the Oriented Malmquist Productivity Index

Nishimizu and Page (1982) were the first to introduce the decomposition of the productivity index into efficiency change and technical change in continuous time and then apply the decomposition in discrete time.[23] Färe et al. (1990, 1992, 1994a) adapted the decomposition to using a non-parametric frontier production function for estimating the efficiency scores. A quest for finding the sources of productivity change followed. I will return to some of these efforts after reviewing the decomposition of Nishimizu and Page (1982) that seems to be overlooked. They were aware of the problems with interpretation in the discrete case:

> Clearly, technological progress and technical efficiency change are not neatly separable either in theory or in practice. In our methodological approach […] we define technological progress as the movement of the best practice or frontier production over time. We then refer to all other productivity change as technical efficiency change. The distinction which we have adopted is therefore somewhat artificial, […]. (Nishimizu and Page (1982), pp. 932–933)

Their approach is set out in Fig. 6.3 (the original Fig. 1, p. 924). All variables are measured in logarithms, and the frontier functions are linear C–D functions with Hicks-neutral technical change from period 1 to period 2. Production is x and input z. The observation A has a production function with the same parameter as the frontiers g_1 and g_2, but with a different constant term. It is then the case that if unit A in period 1 had had the input of period 2, its production level would be at point B. From this point the frontier gap bc is added ending in point C', so $BC' = bc$.

[23] Nishimizu and Page (1982) were the first to refer to a working paper (Caves et al. 1981) that was published as Caves et al. (1982). However, they did not themselves use the term Malmquist productivity index.

Now, the observation in period 2 is found at C greater than C′. Nishimizu and Page then assume that the full potential frontier shift is realised in period 2, but in addition there is a positive efficiency change equal to *C′C*. So, measured in logarithms the productivity change is the sum of the efficiency gap C′C and the frontier gap *BC′* (=*bc*).

Figure 6.4 provides an explanation of their approach in the usual setting of quantities of variables in the simple case of single output and input and the frontiers being CRS. I will now show that the Nishimizu and Page decomposition is the same as the decomposition introduced in Färe et al. (1990, 1992, 1994a, c). A unit is observed at *b* in period 1 and at *f* in period 2. Using the frontier 1 as the benchmark technology instead of the pooled data for all years for simplicity of comparison the Malmquist productivity index (6.14b) for a unit *i* for change between period 1 and 2 and its decomposition are:

$$M_i^1(1,2) = \frac{E^1(y_i^2, x_i^2)}{E^1(y_i^1, x_i^1)} = \frac{E^2(y_i^2, x_i^2)}{E^1(y_i^1, x_i^1)} \cdot \frac{E^1(y_i^2, x_i^2)}{E^2(y_i^2, x_i^2)} = MC_i \cdot MF_i,$$
$$\frac{df/de}{ab/ac} = \frac{df/dg}{ab/ac} \cdot \frac{df/de}{df/dg}, \; MF = \frac{df/de}{df/dg} = \frac{dg}{de} \tag{6.15}$$

The general definition of the Malmquist productivity-change index after the first equality sign is the ratio of the period efficiency measures against the same frontier technology, here for period 1. The expression after the second equality sign shows the multiplicative decomposition into a catching-up[24] measure *MC* and a frontier shift measure *MF*. The second line relates the observations *b* and *f* in Fig. 6.4 to the decomposition in the case of a single output and input. To obtain the correct homogeneity properties we have to use period frontiers that exhibit CRS. We are after information on sources for changes in the Malmquist productivity index, so even if the true contemporary frontier is VRS this does not mean that this frontier is the relevant one to use for the decomposition. I will return to this in the next subsection.

The *MF*-measure represents the relative gap between technologies and is thus the *potential* maximal contribution of new technology to productivity change, while the *MC*-measure is residually determined and may not represent the real efficiency contribution to productivity change, but illustrates the catching-up that is also influenced by the technology shift. It should be observed that the decomposition terms are *multiplied* to give the Malmquist index and not added.

Given that the only information we have about productivity change is the movement of an observation in input—output space, to distinguish between efficiency and technical change is rather difficult. The split into efficiency change and frontier shift that Nishimizu and Page proposed, is, concerning *MF*, based on assuming that the full productivity potential of the frontier shift is actually realised. If both observations had been on their respective frontiers it is obvious that the

[24]To the best of my knowledge this term was first used in Førsund (1993), and then in Fare et al. (1994c).

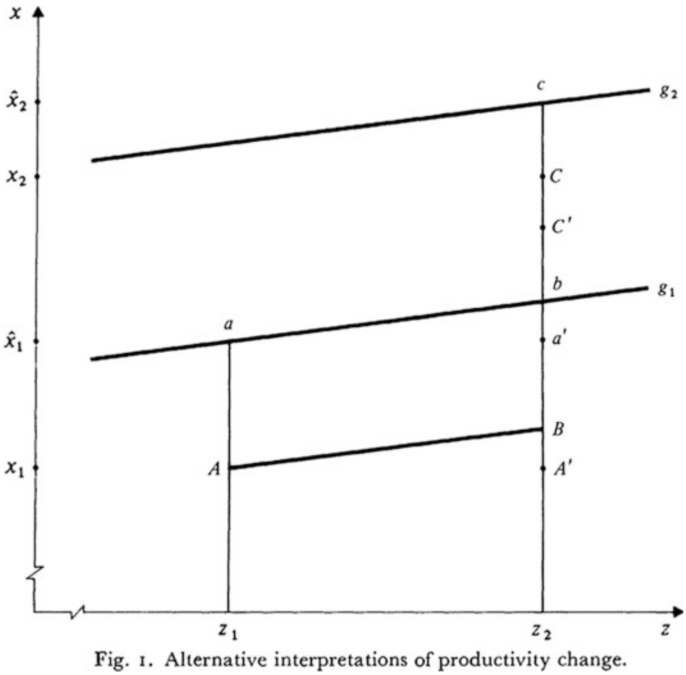

Fig. 6.3 The Nishimizu and Page (1982) decomposition. *Source* The Economic Journal

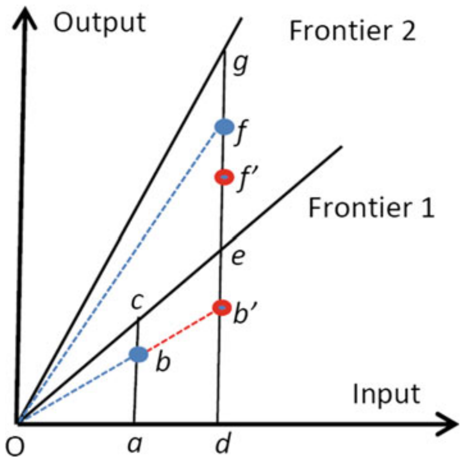

Fig. 6.4 The decomposition of the Malmquist index

Malmquist productivity change will reflect the frontier shift only. If both observations are inefficient with respect to their period frontiers then the efficiency contribution is measured by changing (expanding in Fig. 6.4) the input *Oa* in period

1 to that of *Od* in period 2, but using the actual production function in use in period 1 to predict the hypothetical output level at *f′*. However, I do not operate with any production function for an inefficient observation as Nishimizu and Page did (a CRS C–D function with the same form as the frontier functions), but I will equivalently assume that the efficiency level stays constant getting the inputs of period 2 in period 1. The unit then moves from point *b* to point *b′*. The problem is now to predict where observation *b′* in period 2 will be if the whole potential shift is realised as productivity change. Nishimizu and Page operated with logarithms of the variables and could more conveniently illustrate this, as shown in Fig. 6.3 above. In our Fig. 6.4 this means that the predicted output at point *f′* must obey $df'/db' = dg/de$, the latter being the relative frontier gap Then the same measure for efficiency "contribution" is actually obtained as in Nishimizu and Page, equal to the ratio of the two period efficiency measures. This decomposition is the same as the decomposition introduced in Färe et al. (1990, 1992, 1994a, c). This can be demonstrated in Fig. 6.4 by identifying the efficiency gap as df/df' and the frontier gap df'/db' building on Fig. 1 in Nishimizu and Page (Fig. 6.3 here), and using $df'/db' = dg/de$ and $db'/de = ab/ac$:

$$\frac{df}{df'} \cdot \frac{df'}{db'} = \frac{df/dg}{db'/de} \cdot \frac{dg}{de} = \frac{df/dg}{ab/ac} \cdot \frac{dg}{de} = \frac{df/de}{ab/ac} = M \tag{6.16}$$

However, note that the decomposition does not mean that there is a causation; we cannot distinguish between productivity change due to increase in efficiency and due to shift in technology using the general components in (6.15), as may seem to be believed in some of the empirical literature. The actual productivity change that we estimate using the Malmquist productivity index is from the observation in one period to an observation in another period (from *b* to *f* in Fig. 6.4). The causation is another question related to the dynamics of technical change and how this potential is utilised. As expressed in Nishimizu and Page (1982) after identifying technological progress as the change in the best practice production frontier:

> We then refer to all other productivity change – for example learning by doing, diffusion of new knowledge, improved managerial practice as well as short run adjustment to shocks external to the enterprise – as technical efficiency change. Nishimizu and Page (1982, p. 921)

Nishimizu and Page consider that dynamic factors influence efficiency change, but do not consider the same for realising the new technology.

We cannot decompose efficiency effects and frontier shift effects without making assumptions, according to Nishimizu and Page. Catching up seems to be the best descriptive term for the efficiency component. The decomposition can then be described as the relative potential contribution from technical change multiplied by an efficiency correction factor.

6.3.3 Circularity and Decomposition

Maintaining circularity for both components *MC* and *MF* in the decomposition implies that the technology shift term *MF* will be more complicated. Efficiency measures calculated relative to the benchmark frontier must be involved in the frontier shift measure. A decomposition of the index in Eq. (6.14b) that functions is:

$$M_i^s(u,v) = \underbrace{\frac{E^v(x_{iv},y_{iv})}{E^u(x_{iu},y_{iu})}}_{MC} \cdot \underbrace{\frac{E^s(x_{iv},y_{iv})/E^v(x_{iv},y_{iv})}{E^s(x_{iu},y_{iu})/E^u(x_{iu},y_{iu})}}_{MF},\ i=1,\ldots,N,\ u,v=1,\ldots,T,\ u<v \tag{6.17}$$

The *MF* measure of technology shift is calculated as a 'double' relative measure where both period efficiency measures are relative to the benchmark efficiency measures (Berg et al. 1992). It is easy to see that the decomposition reduces to the Malmquist index (6.14b) by cancelling terms. Notice that to do the decomposition we need benchmark envelopments for each period to be compared in addition to the fixed benchmark envelopment as seen in Fig. 6.2.

It can be illustrated in the case of one output and one input that the frontier shift component still measure the gap between the two benchmark technologies 1 and 2 in Figs. 6.2 and 6.4. Introducing the intertemporal benchmark *s* in Fig. 6.4 we can express the Malmquist index and its components in Fig. 6.5. The observations in period 1 and 2 are marked with blue circles at *b* and *h*. The relative frontier gap between frontier 1 and 2 measured using the observation for period 2 is *fk*/*fg*. We shall see if the decomposition in (6.17) gives the same measure using the notation in Fig. 6.5:

$$M = \frac{fh/fm}{ab/ae} = \underbrace{\frac{fh/fk}{ab/ac}}_{MC} \cdot \underbrace{\frac{(fh/fm)/(fh/fk)}{(ab/ae)/(ab/ac)}}_{MF} \tag{6.18}$$

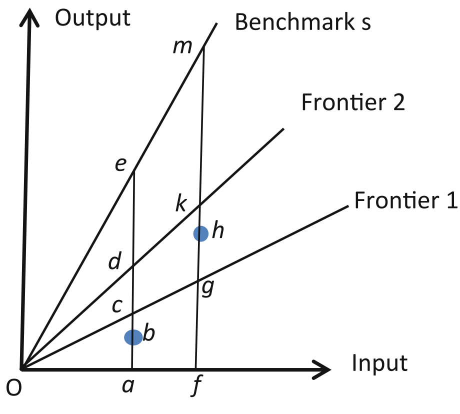

Fig. 6.5 The decomposition of the Malmquist index imposing circularity

The *MF* component can be developed as follows:

$$MF = \frac{(fh/fm)/(fh/fk)}{(ab/ae)/(ab/ac)} = \frac{fk/fm}{ac/ae} \tag{6.19}$$

The last expression is the gap between frontier 2 and benchmark *s* in the numerator and the gap between frontier 1 and the benchmark in the denominator, both expressed as the inverse of the definition of the gap as expressed in the last equation in (6.15). But using the property of like triangles we have $ac/ae = fg/fm$. The last expression in (6.19) can then be written:

$$\frac{fk/fm}{ac/ae} = \frac{fk/fm}{fg/fm} = \frac{fk}{fg} \tag{6.20}$$

This is the relative gap between frontier 2 and 1 using the input for period 2 as the base for calculating the gap.

However, note that in the general multi-output—multi-input case we cannot invoke the property of like triangles; the relative gaps depend on the input and output mixes.

6.3.4 Comments on Decompositions

In Färe et al. (1994b, c) the decomposition into catching up and frontier shift in Färe et al. 1990, 1992, 1994a)[25] was extended to a further decomposition of the efficiency change term into a scale efficiency term and a technical efficiency term, assuming the two contemporaneous frontiers to be VRS. This approach was criticised in Ray and Desli (1997) and a reply given in Färe et al. (1997). In his extensive review of decompositions of the Malmquist index Lovell (2003, p. 442) states: "I conclude that the Färe et al. (1994c) decomposition of the Malmquist productivity index is inadequate".

However, there are problems with the extended decompositions that are not observed by any of the papers above. The first comment is that decompositions are meant to identify sources of impacts on the total factor productivity index of observed movements of a unit in input-output space. It is then not necessarily the

[25]The history of the DEA-based Malmquist productivity index is presented in Färe et al. (1998), Grosskopf (2003) and Färe et al. (2008). The first working paper that established an estimation procedure based on DEA was published in 1989, was presented at a conference in Austin in the same year, and appeared as Färe et al. (1994a); a book chapter in a volume containing many of the conference presentations. The first journal publication appeared as Färe et al. (1990) with an application to electricity distribution. (However, this paper is not referred to in the 2003 and 2008 reviews and neither in Färe et al. (1992), although the methodological approach in the latter is the same).

case that one should use the actual contemporaneous technologies as a point of departure. A point that is under-communicated is the role of the benchmark envelopment. If we want the productivity change index to have the fundamental property of proportionality, then this envelopment have to exhibit constant returns to scale (CRS) even though the true contemporaneous technology is variable returns to scale. It follows most naturally that the decompositions should then also be based on envelopments exhibiting CRS. Thus, I support the choice in Färe et al. (1994c) of a cone benchmark envelopment. Ray and Desli (1997) do not seem to understand this choice, and in Balk (2001, p. 172) it is stated "it is not at all clear why technical change should be measured with respect to the cone technology" in spite of introducing proportionality in his Eq. (2).

Figure 6.2 illustrates the situation; the true contemporaneous technologies may be the variable returns to scale (VRS) functions for the two years, while the benchmark envelopment is represented by the cone CRS(s) based on the pooled data. Now, the catching-up term is the relative distance to the cone envelopments of the data from the two periods, while the frontier shift component is the "double relativity" format of (6.17) also involving distances from the observations to the benchmark envelopment of the pooled data.

There are many followers of the extended multiplicative decomposition in Färe et al. (1994b, c) of decomposing the catching-up term into what is called "pure" technical efficiency and scale efficiency. Pure efficiency is defined as the efficiency of the observation relative to the VRS frontier termed E_2 in Sect. 6.2. Using the terms there we have $E_3 = E_2 \cdot E_5$.[26] The complete decomposition of the change in the catching-up term, assuming a VRS technology for periods u and v and simplifying the notation, dropping writing the variables and unit index, is then

$$\frac{E_{3v}^{v}}{E_{3u}^{u}} = \frac{E_{2v}^{v}}{E_{2u}^{u}} \cdot \frac{E_{5v}^{v}}{E_{5u}^{u}} \tag{6.21}$$

However, it is difficult to see that this decomposition is helpful in interpreting the catch-up term. It is difficult to consider this term as a "driving factor". The E_2 terms are just there to satisfy the identity. The period VRS frontiers do not contribute to the understanding of the productivity changes based on CRS benchmark envelopments constructed by the analyst focusing on the development of the maximal productivity over time. The catch-up term is based on the change in the optimal scale (TOPS). Scale inefficiency has no role in our measure of productivity change. As remarked by Kuosmanen and Sipiläinen (2009, p. 140) "the distinction between the technical change and scale efficiency components is generally ambiguous and debatable." In Balk (2001) change in input mix is also identified as a separate factor, cf. O'Donnell (2012) also including change in output mix. However, these factors are not considered here.

[26]As a control, inserting the definition of E_5 we have for each period technology $E_3 = E_2 \cdot E_3/E_2 = E_3$.

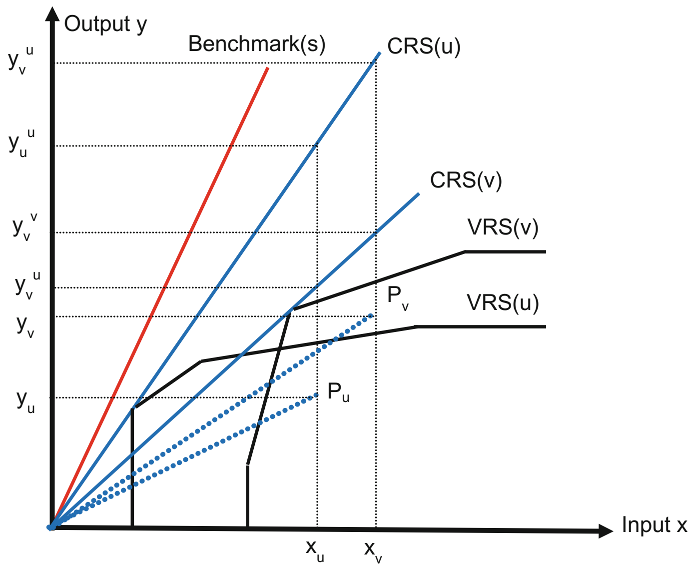

Fig. 6.6 Contemporaneous cones and VRS technologies

From a computational point of view the Malmquist index (6.14b) connects data points with a benchmark envelopment that serves our purpose of measuring productivity change. Pooling the data secures the highest possible degrees of freedom. Decomposition requires the estimation of two additional contemporaneous benchmark envelopments, and reduces the degrees of freedom in the estimation of these in addition to not giving us that much information of real sources behind productivity change.

We may face trouble also with basing the decomposition terms on cone envelopments if estimations of period functions are not properly restricted. An example is given in Fig. 6.6. The VRS envelopments are changed from those in Fig. 6.2 and are crossing each other,[27] and in such a way that the productivity of the optimal scale in period u is greater than in the later period v. We see clearly that the productivity growth measured by the Malmquist index (6.14b) shows growth, but that the frontier shift between periods u, v will show technical regress ($MF < 1$). However, the catching-up component then has to be greater than 1, and so much greater that growth is shown by the product of the terms. Looking at the VRS frontiers where the observations are located conveys that there has been a positive shift in the frontier from period u to period v, but this is the opposite of what the change in the period CRS benchmark tells us. One way to avoid this situation is to use sequential period envelopments. Then the CRS envelopment for period u may

[27]Crossing of technologies and crossing of isoquants as illustrated in Førsund (1993) will be difficult to interpret using geometric means of an index of the type in (6.14a).

be the same as for period v in Fig. 6.6 and productivity growth will be measured as due to efficiency improvement only.

6.4 Conclusions

Efficiency and productivity are two different concepts, but related through the fundamental definition of efficiency as being the relative relationship between the observed productivity of a unit and the maximal achievable productivity for the type of activity in question. Charnes et al. (1978) set up a different route to calculate the same efficiency measures introduced by Farrell (1957) by setting up a ratio form of productivity measures for estimating the efficiency scores, where the weights in the linear aggregation of outputs and inputs are estimated when maximising weighted outputs on weighted inputs subject to no productivity ratio using these weights for all units being greater than one (as a normalisation). However, this way of defining efficiency measures using expressions formally equal to productivity, is not as satisfactory for economists as the Farrell approach, introducing explicitly a frontier production function as a reference for efficiency measure definitions and calculations.

The original Farrell measures developed for constant returns to scale (CRS) has been extended to five efficiency measures for a frontier production function exhibiting variable returns to scale (VRS); input- and output technical efficiency, input- and output scale efficiency, and the technical productivity measure. The relationship between the two measures of technical efficiency involves the average scale elasticity value between the two frontier projection points along the frontier surface. The technical productivity measure and the two scale efficiency measures are developed based on the Frisch (1965) concept of technically optimal scale, predating the use of the concept most productive scale size in the DEA literature with almost 20 years.

It does not seem to be commonly recognised in the DEA literature that in the general case of multiple outputs and inputs the Farrell efficiency measures can all be given productivity interpretations in a more satisfactory way than the ratio form of Charnes et al. (1978). Using quite general theoretical aggregation functions for outputs and inputs with standard properties, it has been shown that all five Farrell efficiency measures can be given a productivity interpretation employing a proper definition of productivity. Each of the two technical efficiency measures and the technical productivity measure can be interpreted as the ratio of the productivity of an inefficient observation and the productivity of its projection point on the frontier, using the general aggregation equations. Of course, we have not estimated any productivity index as such, this remains unknown, but that was not the motivation of the exercise in the first place.

The Malmquist productivity index for bilateral comparisons, applied to discrete volume data and no prices, utilises Farrell efficiency measures directly. In order to have the required index property of proportionality it is sufficient to have as a benchmark an envelopment that exhibits global constant returns to scale, although the underlying contemporaneous production frontiers may have variable returns to scale. One way of obtaining the proportionality properties is basing the benchmark envelopment on the technically optimal scale of the underlying frontiers. If circularity is wanted then this may be done by using cone envelopment for a single year, or pooling all data and using an intertemporal benchmark as is followed in this paper.

Fundamental drivers of productivity change are improvement in efficiency and technical change. The question is how to identify these drivers for a given dataset of outputs and inputs for units. The seminal contribution in Nishimizu and Page (1982) showed one way decomposing a productivity index into a component expressing efficiency change and a component showing the frontier shift impact on productivity that is shown to be the same type of decomposition as the one done for the Malmquist index of productivity change in Färe et al. (1994a). However, a warning of not attaching causality to the decomposition is in place. The decomposition is based on assuming that the full potential of productivity change due to new technology is actually realised, and then the efficiency component is determined residually, but neatly expressed as the relative catching-up to the last period frontier compared with the relative distance to the frontier in the previous period.

If a total factor productivity change index is wanted it is shown that a cone benchmark envelopment satisfy the proportionality test and furthermore using a fixed benchmark technology, for instance based on the pooled dataset as done in this chapter, will satisfy the circularity test. Furthermore, it is argued that cone benchmark envelopments should also be used for contemporaneous frontiers, thus criticising efforts to do further decompositions involving scale efficiencies based on assuming variable returns to scale period frontiers.

References

Atkinson AB, Stiglitz JE (1969) A new view of technological change. Econ J 79(315):573–578

Balk BM (1998) Industrial price, quantity, and productivity indices: the micro-economic theory and an application. Kluwer Academic Publishers, Boston-Dordrecht-London

Balk BM (2001) Scale efficiency and productivity change. J Prod Anal 15(3):159–183

Balk BM, Althin R (1996) A new, transitive productivity index index. J Prod Anal 7(1):19–27

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiency in data envelopment analysis. Manage Sci 30(9):1078–1092

Baumol WJ, Panzar JC, Willig RD (1982) Contestable markets and the theory of industry structure. Harcourt Brace Jovanovich, New York

Belu C (2015) Are distance measures effective at measuring efficiency? DEA meets the vintage model. J Prod Anal 43(3):237–248

Berg SA, Førsund FR, Jansen ES (1992) Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. Scand J Econ 94(Supplement):S211–S228

Bjurek H (1996) The Malmquist total factor productivity index. Scand J Econ 98(2):303–313

Bjurek H, Førsund FR, Hjalmarsson L (1998) Malmquist productivity indexes: an empirical comparison. In Färe R, Grosskopf S, Russell RR (eds) Index numbers: essays in honour of Sten Malmquist. Kluwer Academic Publishers, Boston, Essay 5, pp 217–239

Brännlund R, Lundgren T (2009) Environmental policy without cost? A review of the Porter hypothesis. Int Rev Environ Resour Econ 3:75–117

Caves DW, Christensen LR, Diewert E (1981) A new approach to index number theory and the measurement of input, output, and productivity. Discussion Paper 8112, Social Systems Research Institute, University of Wisconsin

Caves DW, Christensen LR, Diewert E (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50(6):1393–1414

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Cornwell C, Schmidt P, Sickles R (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econometrics 46(1–2):185–200

Diewert WE (1992) Fisher ideal output, input and productivity indexes revisited. J Prod Anal 3 (3):211–248

Färe R, Grosskopf S, Yaisawarng S, Li SK, Wang Z (1990) Productivity growth in Illinois electric utilities. Resour Energy 12:383–398

Färe R, Grosskopf S, Lindgren B, Roos P (1992) Productivity changes in Swedish pharmacies 1980-1989: a non-parametric Malmquist approach. J Prod Anal 3(1/2):85–101

Färe R, Grosskopf S, Lovell CAK (1994b) Production frontiers. Cambridge University Press, Cambridge

Färe R, Grosskopf S, Lindgren B, Roos P (1994a) Productivity developments in Swedish hospitals: a Malmquist output index approach: In Charnes A, Cooper WW, Lewin AY, Seiford LM (eds) Data envelopment analysis: theory, methodology, and applications. Kluwer Academic Publishers, Boston/Dordrecht/London, Chapter 13, pp 253–272. (Also published in 1989 as *Discussion paper 89–3*, Department of Economics, Southern Illinois University, Carbondale)

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

Färe R, Grosskopf S, Norris M (1997) Productivity growth, technical progress, and efficiency change in industrialized countries: reply. Am Econ Rev 87(5):1040–1044

Färe R, Grosskopf S, Roos P (1998) Malmquist productivity indexes: a survey of theory and practice. In Färe R, Grosskopf S, Russell RR (eds) Index numbers: essays in honour of Sten Malmquist. Kluwer Academic Publishers, Boston/London/Dordrecht, Essay 3, pp 127–190

Färe R, Grosskopf S, Margaritis D (2008) Efficiency and productivity: Malmquist and more. In Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency and productivity growth. Oxford University Press, New York, Chapter 5, pp 522–622

Farrell MJ (1957) The measurement of productive efficiency. J Roy Stat Soc Ser A (General) 120 (3):253–281(290)

Farrell MJ, Fieldhouse M (1962) Estimating efficient production functions under increasing returns to scale. J Roy Stat Soc Ser A (General) 125(2):252–267

Førsund FR (1992) A comparison of parametric and non-parametric efficiency measures: the case of Norwegian ferries. J Prod Anal 3(1):25–43

Førsund FR (1993) Productivity growth in Norwegian ferries. In Fried HO, Lovell CAK, Schmidt SS (eds) The measurement of productive efficiency. Oxford University Press, Oxford, Chapter 14, pp 352–373

Førsund FR (1996) On the calculation of the scale elasticity in DEA models. J Prod Anal 7 (3):283–302

Førsund FR (1998) The rise and fall of slacks: comments on Quasi-Malmquist productivity indices. J Prod Anal 10(1):21–34

Førsund FR (2010) Dynamic efficiency measurement. Indian Econ Rev 45(2):125–159

Førsund FR (2015) Economic perspectives on DEA. Memorandum No 10/2015 from the Department of Economics, University of Oslo

Førsund FR, Hjalmarsson L (1974) On the measurement of productive efficiency. Swed J Econ 76 (2):141–154

Førsund FR, Hjalmarsson L (1979) Generalised Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. Econ J 89(354):294–315

Førsund FR, Hjalmarsson L (2004a) Are all scales optimal in DEA? Theory and empirical evidence. J Prod Anal 21(1):25–48

Førsund FR, Hjalmarsson L (2004b) Calculating scale elasticity in DEA models. J Oper Res Soc 55(10):1023–1038. doi:10.1057/palgrave.jors.2601741

Førsund FR, Hjalmarsson L, Krivonozhko VE, Utkin OB (2007) Calculation of scale elasticities in DEA models: direct and indirect approaches. J Prod Anal 28(1):45–56

Førsund FR, Kittelsen SAC, Krivonozhko VE (2009) Farrell revisited–visualizing properties of DEA production frontiers. J Oper Res Soc 60(11):1535–1545

Fried HO, Lovell CAK, Schmidt SS (eds) (2008) The measurement of productive efficiency and productivity growth. Oxford University Press, New York

Frisch R (1965) Theory of production. D. Reidel, Dordrecht

Gini C (1931) On the circular test of index numbers. Metron 9(2):3–24

Grifell-Tatjé E, Lovell CAK (1995) A note on the Malmquist productivity index. Econ Lett 47 (2):169–175

Grosskopf S (2003) Some remarks on productivity and its decompositions. J Prod Anal 20(3):459–474

Hanoch G (1970) Homotheticity in joint production. J Econ Theor 2(4):423–426

Krivonozhko V, Volodin AV, Sablin IA, Patrin M (2004) Constructions of economic functions and calculation of marginal rates in DEA using parametric optimization methods. J Oper Res Soc 55(10):1049–1058

Kuosmanen T, Sipiläinen T (2009) Exact decomposition of the Fisher ideal total factor productivity index. J Prod Anal 31(3):137–150

Lovell CAK (2003) The decomposition of Malmquist productivity indexes. J Prod Anal 20 (3):437–458

Nishimizu M, Page JM (1982) Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia 1965-78. Econ J 92(368):920–936

O’Donnell CJ (2012) An aggregate quantity framework for measuring and decomposing productivity change. J Prod Anal 38(3):255–272. doi:10.1007/s11123-012-0275-1

Panzar JC, Willig RD (1977) Economies of scale in multi-output production. Quart J Econ 91 (3):481–493

Pastor JT, Lovell CAK (2005) A global Malmquist productivity index. Econ Lett 88(2):266–271

Podinovski VV, Førsund FR (2010) Differential characteristics of efficient frontiers in data envelopment analysis. Oper Res 58(6):1743–1754

Podinovski VV, Førsund FR, Krivonozhko VE (2009) A simple derivation of scale elasticity in data envelopment analysis. Eur J Oper Res 197(1):149–153

Porter ME, van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. J Econ Perspect 9(4):97–118

Ray SC, Desli E (1997) Productivity growth, technical progress and efficiency change in industrialized countries: comment. Am Econ Rev 87(5):1033–1039

Samuelson PA, Swamy S (1974) Invariant economic index numbers and canonical duality: survey and synthesis. Am Econ Rev 64(4):566–593

Shephard RW (1970) Theory of cost and production functions (first edition 1953). Princeton University Press, New Jersey

Starrett DA (1977) Measuring returns to scale in the aggregate, and the scale effect of public goods. Econometrica 45(6):1439–1455

Tulkens H, van den Eeckaut P (1995) Non-parametric efficiency, progress, and regress measures for panel data: methodological aspects. Eur J Oper Res 80(3):474–499