MEMORANDUM

No 27/2008

Dynamic Analysis of Structural Change and Productivity Measurement

Finn R. Førsund
Lennart Hjalmarsson

Department of Economics
University of Oslo

Last 10 Memoranda

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

DYNAMIC ANALYSIS OF STRUCTURAL CHANGE AND PRODUCTIVITY MEASUREMENT[*]

by

Finn R. Førsund

Department of Economics, University of Oslo

Lennart Hjalmarsson,

Department of Economics, University of Gothenburg

October 2008

Abstract: A dynamic analysis of structural change is reviewed based on a vintage model of substitutability between inputs including capital before investment, but no substitution possibilities after investment, and ex post production possibilities characterised by fixed input coefficients. Key elements in understanding structural change are the entering of capacity embodying new technology and exiting of capacity no longer able to yield positive quasi-rent. Three production function concepts are identified: the ex ante micro unit production function as relevant when investing in new capacity, the ex post micro production function, and the short-run industry production function giving the production possibilities at the industry level. Productivity measurement, taking these types of production functions into consideration, leads to different interpretations of productivity change than traditional approaches not being clear about which production function concept is used.

---

1. Introduction

There is an obvious link between economic growth and the performance of industries within an economy. Performance of economies is, in fact, often measured by productivity growth, as remarked in Diewert and Fox (2008). Productivity change can be measured at various levels of aggregation; the national level, the industry level and the level of individual firms, or even a finer disaggregation. Both between industries and within an industry we have reallocation of resources between different industries and different firms during economic growth, constituting a dynamic ongoing process. We will focus on the level of an industry and the firms within an industry. The approach followed is a bottom-up approach from firm level in order to capture the range of dynamic changes, and based on modelling technology compatible with engineering information, implying that the firms must be fairly homogeneous as to types of outputs and inputs.

Analyses of structural change and productivity growth have a long tradition in the economics literature. Marshall (1890) was interested in the connection between the short- and the long run, and used the picture of the life cycle of a representative unit; it grows and declines like a tree in a forest. However, using the device of a representative unit risks missing an important insight about dynamic structural change: the structure of trees is represented by the forest, and not by the life cycle of a single tree. (Wedervang (1964) focuses on the population of firms).

Empirical economists have represented the structure of an industry in a simple way by comparing average practice, measured by single factor productivity, especially labour, but also energy, with best practice within the industry (Mitchell, 1937; Åkerman, 1931). Future use of inputs when average practice catches up with best practice can then be predicted, as well as the time lag between best practice and average practice (Maywald, 1957).

A significant extension of the structural analysis was done in Salter (1960), and he also studied the dynamic impact of structural change on productivity change. He incorporated (without realising this) a much earlier method introduced by Heckscher (1918).

An explicit micro foundation of the Salter dynamic analysis was accomplished in Johansen (1972), formalising the vintage or putty-clay model for micro units. He stressed the distinction between available blueprints of the most recent technology when investing in a production unit, and the current production possibilities of an industry comprising units producing the same output. The purpose of the present paper is to show implication for measurement of productivity change for an industry when the vintage approach of Johansen is the relevant model. As an example of how to follow up a study in the Johansen spirit we draw upon a study of the Norwegian aluminium industry (Førsund and Jansen, 1983).

The outline of the paper is as follows. In Section 2 the dynamic structural analysis of Salter is reviewed, and in Section 3 the Johansen vintage model is established, and the investment decision of a single firm is worked through. Section 4 treats in detail the Johansen short-run industry production function and uses data from the Norwegian aluminium industry 1966 – 1978 to make empirical analyses of productivity changes. In Section 5 there is a discussion of how to measure technical advance of the ex ante micro-unit function, and the relationship between some recently developed ways of measuring productivity change using the Malmquist index and the vintage approach of Johansen. Section 6 concludes.

2. Dynamic structural analysis

A first step in an empirical analysis is to sort all firms in an industry according to increasing variable unit cost, thus creating a *merit-order* sorting of the firms. The difference between the going market price and the unit costs shows what is available for remuneration of fixed capital. This is the *quasi-rent*, introduced by Marshall (1890). If the quasi-rent is negative, then this firm is a candidate for being closed down. Even with a positive quasi-rent a firm may be unable to service borrowed financial capital and ownership may change hands, but it is in general profitable in a social sense to keep producing as long as the quasi-rent is positive.

Heckscher (1918) used a simple sorting of firms according to unit costs as the point of departure of his analysis of consequences for domestic firms of reducing the duty on imports of a competing industry's product and hence a lowering of the market price. We therefore, in
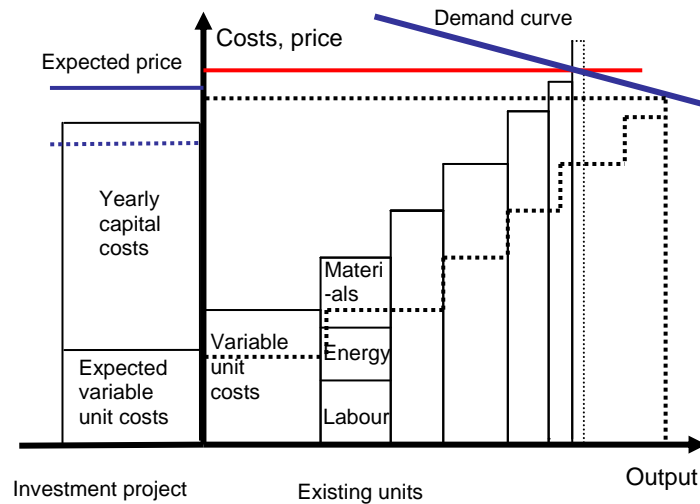
Figure 1. Salter's dynamic analysis

Førsund and Hjalmarsson (1987), termed the diagram showing sorted unit costs and the market price line as a *Heckscher diagram*. This diagram is illustrated in the right-hand part of Figure 1. The sorting after increasing unit costs make the distribution into a merit-order curve. The output capacity of firms is proportional to the length of the unit cost histograms. We see that there is one small unit earning negative quasi-rent at the prevailing market price. This firm is assumed not to be operating, indicated by the dashed lines. Firms with different cost characteristics and representing different technologies and reflecting different more or less outdated blueprint technologies, coexists and earn positive quasi-rents.

Subdividing the variable unit cost into components such as labour, energy and materials, as done for the second unit in the Heckscher diagram in Figure 1, Wohlin (1970) could discuss the impact on total unit costs of different increases in the costs of types of inputs. The consequences for, e.g., the aluminium industry of an increase of the electricity price can be analysed using a Heckscher diagram with a subdivision of costs (Bye et al., 2006).

Salter (1960, p. 59) extended the Heckscher diagram by introducing a potential investment in production capacity, exhibiting best practice techniques.[1] The new technology may imply lower unit variable cost than the most efficient existing unit, but in addition to variable costs capital cost must now be considered when making an investment decision. The situation is set

---

[1] The dynamic analysis of Salter is termed the *Salter cycle* in Ayres (1999).

out in the left part of Figure 1. An investment with the capacity represented by the width of the histogram is profitable if the output price the investor considers is high enough to cover both variable and capital costs, the latter expressed as yearly costs. The solid price line indicates a profitable investment, while the dotted price line represents a price level that makes the investment unprofitable.

Assuming that the indicated investment is undertaken, the new unit-cost distribution shifts to the right and is shown by the dotted lines of the step curve in the Heckscher diagram of Figure 1. The dotted lines represent the new merit-order cost curve. The reaction in the market to the entry of new capacity is shown by the new, lower short-run equilibrium price (horizontal dotted line) as the intersection of the supply curve of firms earning positive quasi-rent and the demand curve. As a result of new capacity one firm runs into negative quasi-rent, and is removed from the supply curve. The reduction (maybe unexpected for the investors) in the market price due to investment in capacity may be typical for a capital-intensive industry that is uncoordinated as to investment decisions. The naive prospects for the investments looked better ex ante than the ex post reality.

3. The putty – clay model

The dynamic structural analysis of Salter (1960) can be founded on an operational production theory for firms. Such a production theory was presented in Johansen (1972). The point of departure there is that the nature of production function concepts to be estimated is often not stated clearly enough in the empirical literature. "The crudeness of the concept of the production function, as it is being used in most econometric research, is accordingly out of proportion with the sophistication of the theories and methods by which it [the notion of production functions] is surrounded" (Johansen, p.1). Four concepts of production functions are introduced:

   a) The micro unit ex ante production function
   b) The micro unit ex post production function
   c) The industry short-run production function
   d) The industry long-run production function

The ex ante production function for a micro unit (firm, plant, or a piece of equipment) exhibits the standard neoclassical properties of substitutability between inputs, including capital as an input. But once the investment is made the ex post micro production function is characterised by fixed variable input coefficients, and there are no longer any substitution possibilities. The choice of factor proportions made on the ex ante function is "frozen" into fixed variable input coefficients ex post, and capital only serves the role of defining the capacity limit. This is the extreme version of a vintage (or putty-clay, as Phelps (1963) termed it) production function.

The short-run industry production function is a construct of the analyst. The normative question behind it is how the given capacities of the micro units should be utilised. An aggregate production function is introduced utilising existing capacities in a certain way, and it is possible that the actual utilisation of units may deviate from this reference utilisation.

The last production function concept d) covers the situation in steady state with no technical change, and is more of interest in analyses of long-term growth. We will not be concerned with this concept below.

For simplicity we will only specify a single output, but use multiple inputs. The ex ante production function for a micro unit to be constructed can formally be expressed as

$$y = f_o(x_1, ..., x_n, K), f'_{oi}, f'_{oK} > 0, i = 1, ..., n \tag{1}$$

where y is output, x_i ($i = 1, ..., n$) current inputs and K is real capital. Since we are dealing with vintages the variables and the production function must in principle be dated both with current time and the time of the construction of the vintage. In order to simplify, only the ex ante production function is dated with the current time $t = 0$ in (1). The production function is assumed to have standard neoclassical properties as to marginal productivities and exhibit variable returns to scale.

When a choice is made from the ex ante function, i.e., a point on an isoquant is selected, the volume of capital is fixed ex post, and we assume that there is no longer any substitution possibilities between variable inputs. The ratios of input per unit of output are frozen and reflecting the point picked on the isoquant of the ex ante function. Correspondingly, there is an upper limit on the output capacity given by the output label of the chosen isoquant. In

contrast to the literature on quasi-fixed factors the nature of the production function changes fundamentally between the ex ante and the ex post function. The latter can be expressed as an limitational law, or as a Leontief production function:

$$y(t,v) = \min\left[\frac{x_1(t,v)}{\xi_1(v)}, ..., \frac{x_n(t,v)}{\xi_n(v)}, \overline{y}(v)\right], t \geq v \qquad (2)$$

where t is current time and v the time of investment. The output capacity is $\overline{y}(v)$. The "frozen" input coefficients are defined by

$$\xi_i(v) = \frac{\overline{x}_i(t,v)}{\overline{y}(v)} , i = 1,..,n \qquad (3)$$

where $\overline{x}_i(t,v)$ is the full capacity use of variable input i. There is no explicit capital variable in the short-run function, only capacity of the unit measured by output. Capital is indirectly represented by the production capacity. Assuming that there is no waste in the short run implies that all input coefficients can be measured by observed inputs and output. Then no sophisticated estimation method is needed for the coefficients. However, it may be more realistic to assume that the input coefficients vary with the capacity utilisation, e.g., becoming larger the smaller the capacity utilisation, requiring more sophisticated methods than calculating ratios.

One may try to test this extreme version of the vintage model econometrically (Belifante, 1978; Fuss, 1978). One problem will then be the nature of the production unit. A firm may consist of many plants, and a plant may again comprise several distinct types of equipment, and it may be that the extreme vintage assumption is most suitable for the most disaggregated level of a piece of equipment. However, concerning testing the vintage hypothesis it may also be a good idea to follow the recommendation provided by Leif Johansen: "In fact, I think the best way to test the putty-clay hypothesis is simply to inspect production equipment and talk with engineers and technicians" (Johansen, 1972, p. 226).

The investment decision

In order to bring out key characteristics of investment decisions we will make the following simplifying assumptions:

       a) Certainty about future prices

       b) Only one unit consisting of a single vintage

       c) Economic lifetime less than physical lifetime

   d) Investment at time $t = 0$

   e) Disregarding timing of investment

   f) Disregarding that carrying out investment takes time

   g) Full capacity utilisation until scrapping

   h) No scrap value of capital

   i) No disembodied technical change

   j) No maintenance costs of capital

It will be straightforward to generalise some of the assumptions above, while other generalisations require more complex analysis.

Under the assumptions above the present value of net profit over the economic lifetime, T, for the firm at time $t = 0$ may be written

$$\pi(0) = \int_{t=0}^{T} e^{-rt}[p(t)y(t,0) - \sum_{i=1}^{n} q_i(t)x_i(t,0)]dt - q_K(0)K(0) \tag{4}$$

The input prices are q_i (i = 1,..,n) and the price per unit of capital q_K and r is the fixed discount rate. The integral is the present value of the quasi-rent. The firm has to make an initial choice as to output level and levels of variable inputs and capital, resulting in fixed input coefficients for the operation of the firm. By assumption there is full capacity utilisation over the economic lifetime, T, which has to be determined at time $t = 0$. The optimisation problem can be stated as

$$Max \ \pi(0) = \int_{t=0}^{T} e^{-rt}[p(t)y(t,0) - \sum_{i=1}^{n} q_i(t)x_i(t,0)]dt - q_K(0)K(0)$$

subject to $\tag{5}$

$$y(0,0) = f_o(x_1(0,0),...,x_n(0,0), K(0)) \text{ for } t = 0,$$
$$y(t,0) = \bar{y}(0) \ , \ x_i(t,0) = \xi_i(0)\bar{y}(0) \text{ for } t \in [0,T]$$

After determining implicitly the input coefficients for variable inputs, capital investment and maximal output based on the ex ante function, as expressed by the first constraint, the limitational law (2) holds for the actual production, as expressed by the second constraint.

The necessary first-order conditions for interior solutions for the variable inputs are

$$\frac{\partial \pi(0)}{\partial x_i(0,0)} = \int_{t=0}^{T} e^{-rt}[p(t)\frac{\partial f_o}{\partial x_i(0,0)} - q_i(t)]dt = 0 \Rightarrow$$

$$\int_{t=0}^{T} e^{-rt}p(t)dt \frac{\partial f_o}{\partial x_i(0,0)} = \int_{t=0}^{T} e^{-rt}q_i(t)dt \ , i = 1,...,n \tag{6}$$

The first term on the left-hand side of the last equation is the present value of the marginal productivity of variable input i. This value should be set equal to the present value of the outlay on a unit of the input. Current prices in the static textbook case are replaced with the present value of prices. The current value of the marginal productivity may now never be equal to the current value of the input price. This relationship will only hold in an average sense, forming the average of prices dividing the integrals by the economic lifetime, T.

The choice of factor ratios will be directly influenced by the forecasted input prices:

$$\frac{\frac{\partial f_o}{\partial x_i(0,0)}}{\frac{\partial f_o}{\partial x_j(0,0)}} = \frac{\int_{t=0}^{T} e^{-rt} q_i(t)dt}{\int_{t=0}^{T} e^{-rt} q_j(t)dt} \ , \ i,j = 1,...,n \tag{7}$$

A factor with a relative low average price will be substituted for a factor with a relatively high average price. For example, if the wage rate is expected to increase more than the energy price, the initial choice will be to use relatively more energy for a given output level.

The necessary first-order condition for capital is

$$\frac{\partial \pi(0)}{\partial K(0)} = \int_{t=0}^{T} e^{-rt} p(t) \frac{\partial f_o}{\partial K(0)} dt - q_K(0) = 0 \Rightarrow$$

$$\int_{t=0}^{T} e^{-rt} p(t)dt \frac{\partial f_o}{\partial K(0)} = q_K(0) \tag{8}$$

The present value of the marginal productivity is set equal to the capital price. Combining (6) with (8) we have that a project with a given output capacity will be more capital intensive the higher the present value of input prices are relatively to the initial capital price.

Inserting the first-order conditions (6) and (8) into the profit expression (4) yields

$$\pi(0) = \int_{t=0}^{T} e^{-rt}[p(t)y(t,0) - \sum_{i=1}^{n} q_i(t)x_i(t,0)]dt - q_K(0)K(0) =$$

$$\int_{t=0}^{T} e^{-rt} p(t)dt \overline{y}(0) - \sum_{i=1}^{n} \int_{t=0}^{T} e^{-rt} p(t)dt \frac{\partial f_o}{\partial x_i(0,0)} x_i(0,0)dt - \int_{t=0}^{T} e^{-rt} p(t)dt \frac{\partial f_o}{\partial K(0)} K(0) = \tag{9}$$

$$\int_{t=0}^{T} e^{-rt} p(t)dt \overline{y}(0)(1-\varepsilon_o)$$

The last expression is obtained employing the passus equation (Frisch, 1965), where ε_o is the passus coefficient of the ex ante function (1). If it is optimal to chose an output level equal to

the optimal scale output, then the present value of profit is zero, and the rate of return on the investment is equal to the rate of calculation, r. However, it may be optimal to have a level of output greater than optimal scale, and then the rate of return on the capital investment will be greater than the rate of calculation.

The determination of the economic lifetime follows from the condition

$$\frac{\partial \pi(0)}{\partial T} = e^{-rT}[p(T)y(T,0) - \sum_{i=1}^{n} q_i(T)x_i(T,0)] = 0,$$ (10)

implying

$$p(T)y(T,0) = \sum_{i=1}^{n} q_i(T)x_i(T,0) \Rightarrow p(T) = \sum_{i=1}^{n} q_i(T)\frac{x_i(T,0)}{y(T,0)} = \sum_{i=1}^{n} q_i(T)\xi_i(0),$$ (11)

inserting the input coefficients (3). The last right-hand expression of (11) is the variable unit cost of production a time T. Production will be terminated when the current variable unit cost becomes higher than the current output price. The difference between the output price and the variable unit cost is the unit quasi-rent. Thus, the economic lifetime under our assumption is determined as a quasi-rent criterion: production is terminated when the quasi-rent becomes negative.

4. The short-run industry production function

In the short run the total output and use of current inputs in an industry is determined by the utilisation of individual firm output capacities and the short-run micro productions function (2) with the input coefficients determined by (3). In Johansen (1972) a production function covering the industry as a unit was defined, using the classical definition of a production function. The industry consists of N units with homogenous output and inputs. This procedure can be regarded as a special kind of aggregation: the question asked is how given current inputs and the given micro-unit capacities should be utilised in order for the aggregated industry output to be maximised:

$$\text{Max } Y = \sum_{j=1}^{N} y_j$$

subject to

$$\sum_{j=1}^{N} \xi_{ij} y_j \le \overline{X}_i , i = 1,...,n \qquad (12)$$

$$y_j \le \overline{y}_j , j = 1,...,N$$

Total output and total inputs are denoted by uppercase letters. The formulation is built on the short-run micro production functions (2). The solution of problem (12) yields an optimal way to utilise resources, and no prices are involved. The observed way of utilising resources and capacities may deviate from the solution of (12), so what is introduced is a benchmark for optimal utilisation of the micro units given available total resources.

The Lagrangian for the problem is

$$L = \sum_{j=1}^{N} y_j - \sum_{i=1}^{n} \lambda_i (\sum_{j=1}^{N} \xi_{ij} y_j - \overline{X}_i) - \sum_{j=1}^{N} \gamma_j (y_j - \overline{y}_j) \qquad (13)$$

The necessary first-order conditions are

$$\frac{\partial L}{\partial y_j} = 1 - \sum_{i=1}^{n} \lambda_i \xi_{ij} - \gamma_j \le 0 (= 0 \text{ for } y_j > 0)$$

$$\lambda_i \ge 0 (= 0 \text{ for } \sum_{j=1}^{N} \xi_{ij} y_j < \overline{X}_i) \qquad (14)$$

$$\gamma_j \ge 0 (= 0 \text{ for } y_j < \overline{y}_j)$$

An optimal solution implies that a micro unit may be in one of three states: fully utilised, partly utilised, or not used at all. The shadow price, λ_i, on the input constraint i has the interpretation, in an optimal solution, of the change in the objective function of a change in the resource availability, i.e., the shadow price shows directly the marginal productivity of the resource in question. Conditions (14) give the characterisation of the three states:

    1) Fully utilised units: $1 - \sum_{i=1}^{n} \lambda_i \xi_{ij} = \gamma_j \ge 0$

    2) Partly utilised units: $1 - \sum_{i=1}^{n} \lambda_i \xi_{ij} = 0$

    3) Units not in use: $\quad 1 - \sum_{i=1}^{n} \lambda_i \xi_{ij} \le 0$

The common expression on the left-hand sides above resembles the definition of quasi-rent in Section 3. The measurement unit of the shadow price is output per unit of input i, and the

input coefficient is measured as input of type *i* per unit of output of micro unit *j*. The whole expression can be interpreted as the unit quasi-rent deflated with a common output price for the case of all marginal productivities of the micro units being the same. The measurement unit for a factor price deflated by the output price is just output per unit of input, i.e., the same unit as for productivity. In the case of a fully utilised unit the quasi-rent will typically be positive, a partly utilised unit will have zero quasi-rent, and an inactive unit will typically face a negative quasi-rent if positive production is undertaken.

Assuming we have a unique optimal solution the endogenous variables can be written as functions of the exogenous variables:

$$y_j = F_j(\bar{X}_1,...,\bar{X}_n, \bar{y}_1,..,\bar{y}_N)\,(j=1,...,N) \Rightarrow$$
$$Y = F^*(\bar{X}_1,...,\bar{X}_n, \bar{y}_1,..,\bar{y}_N) = F(X_1,...,X_n)$$

(15)

In the last equation we have aggregated the outputs of the micro unit and suppressed capacities as arguments, as well as considering a parametric variation in the given total inputs. The relation $F(.)$ is the short-run industry production function. Notice that capital has no role as an argument in the function. The output capacities of the micro units are arguments in the function, but these capacities are fixed in the short run, and therefore for notational convenience included in the functional form.

In a dynamic perspective the short-run function reflects the history of the ex ante function over time, and the choices of factor ratios made. The standard characterisation of a production function by substitution- and scale properties also apply to the short-run function. However, with the micro functions having fixed input coefficients the isoquants will be piecewise linear. An empirical illustration of the shape of the short-run function and isoquants for the Norwegian aluminium industry is given in Figure 2. The most "efficient" units, i.e., the units with the highest quasi-rent, will be utilised first, and then less efficient units will be taken into use. This kind of pattern of utilisation may be termed merit order, as is typically used for optimal utilisation of, e.g., thermal units in the electricity production sector. The system operator of the electricity-generating sector solves our problem of finding a short-run industry function and uses it to call on generating units according to the conditions (14).[2] We see that the substitution regions are quite narrow, in contrast to the usual exhibition in textbooks. The

---

[2] In practice merit order is based on costs, therefore the theoretical merit order based on quasi-rent will only be relevant if input prices are equal.

narrowness is due to the fact that the aluminium plants are quite similar as to input coefficients. The piecewise linearity of the isoquants is evident.

Productivity measures for the short-run industry production function

It is natural to measure the change in productivity between years by following the change of the isoquants representing the same output level. Some levels are indicated in Figure 2. Following for instance the 300 kt isoquant across the years the visual impression is of a shift to the left, implying a strong labour-saving bias. It is not so easy to see whether the isoquant has shifted towards the origin. Salter (1960) proposed in general to measure productivity change by calculating the change in average costs for the same output level and keeping the input prices fixed. This definition will serve us well also in the setting of the short-run industry production function. Capital as an input is then left out, but in the putty-clay world all capital is a gift from the past, and it should not logically be considered when calculating productivity change within short-run functions. The return on capital investment is another question in the putty-clay model, as discussed in Section 3. As stated above, the shape of the
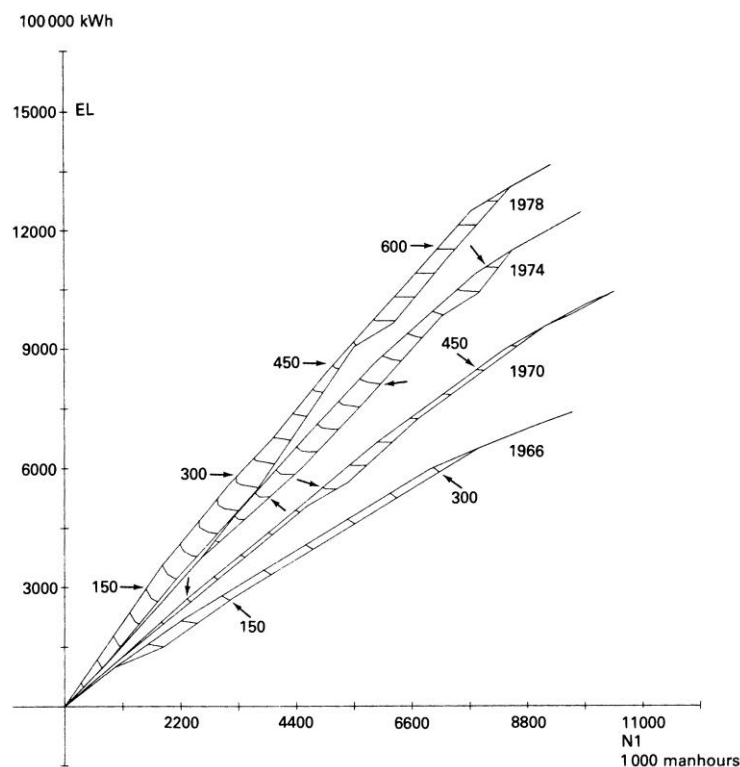
Source: Førsund and Jansen (1983)

Figure 2. Substitution regions and isoquants
for the Norwegian aluminium industry

short-run functions reflects the history of the ex ante functions, and capital is an argument in these functions. When considering sources of productivity growth within the short-run function new capital is one of the sources. We will return to productivity measurement and the ex ante function below.

Salter's proposal for measuring change of best-practice productivity (Salter, 1960, pp. 27-29) may be applied to the short-run industry function by using the cost functions for the short-run function:

$$\frac{c_{t+1}(q_1,...,q_n,Y)/Y}{c_t(q_1,...,q_n,Y)/Y} \tag{16}$$

where $c_t(.)$ is the cost function for the short-run function at t corresponding to the production function (15), and q_i ($i = 1,..,n$) is an imputed input price (using the marginal productivities). Productivity increases if the ratio is less than 1 and decrease if it is greater than 1. The cost function applies to the industry as a unit, and the factor prices are not necessarily the ones faced by each individual unit. However, for an empirical application, by construction we must use the same input prices for all firms. Taking an average of observed prices across firms for a period seems to be the simplest procedure. The prices must be kept fixed, so a question is the choice of base year. If the prices of the first period is chosen this corresponds to a Laspeyres approach, while choosing the prices of the last period corresponds to a Paasche approach.

In Figure 3 both the marginal cost, c_Y, and the average cost, c/Y, are set out based on the last year's average prices. The location of points of intersection of the expansion path within the substitution regions (the latter shown in Figure 2) with the isoquants gives us the factor point for calculating minimum costs for the given output level. For the 300 kt isoquant we see a marked productivity improvement both from 1966 to 1970, and from 1970 to 1978. (The numbers are 28%, respectively 20%, implying average growth rates of 6% and 2%, see Førsund and Jansen, 1983.) This progress is very difficult to see from Figure 2, where changes in factor requirements, i.e., the labour saving bias, catch the attention.

As is well known from the literature, the ratio of the average cost and the marginal cost shows the scale elasticity of the short-run function. By construction the function must have a scale elasticity in the interval (0,1], and the scale elasticity is falling when moving outwards along an expansion path up to the point where all capacity is exhausted. We see that for the 300 kt isoquant the scale elasticity is markedly smaller in 1966 than in the two other years, reflecting
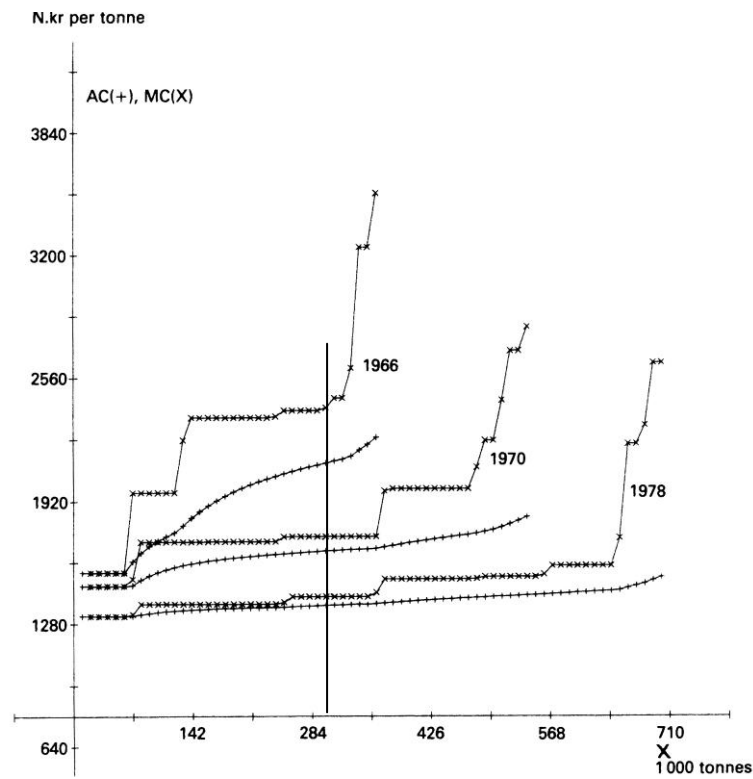
Source: Førsund and Jansen (1983)

Figure 3. Marginal- and average cost functions
for the Norwegian Aluminium industry

the change that has taken place, yielding a more homogeneous structure of the firms at the output level in question.

Bias of technical change

The labour saving bias is evident from Figure 2. Salter proposed to use as bias measure, D_{ij}, the factor ratios between factors i and j, keeping the output level and factor prices constant, i.e., using factor points generated by expansion paths as functions of the same factor prices:

$$D_{ij} = \frac{(X_{i,t+1} / X_{j,t+1})\big|_{Y_o,q_i}}{(X_{i,t} / X_{j,t})\big|_{Y_o,q_i}}, \; i, j = 1,..,n, i \neq j \tag{17}$$

The ratio of electricity to labour increased with 31% from 1966 to 1970, and with 53% from 1970 to 1978, quite in correspondence with the visual impression of Figure 2.

By transforming the substitution regions and the isoquant maps of Figure 2 into input coefficient space we can get another visual impression both of productivity change and of bias

effects. This transformation also allows us to visualise the movement of the best-practice part of the short-run function since the frontier towards the origin is portrayed, as seen in Figure 4. The labour-saving bias is clearly portrayed. It is also remarkable that the best-practice parts are almost stationary as to electricity productivity. This indicates that at best practice the technical change over a 12-year span concerning electricity use has been very modest. The year 1974 is an exception, and closer investigation revealed that this year the capacity utilisation was extraordinarily high, explaining the higher electricity efficiency. The improvement in productivity is due to improvements of firms moving towards the technology frontier as to use of electricity. (A theoretical physical minimum can be calculated.) There is a clear improvement of the least efficient part over time. The reversal between 1974 and 1978 is due to lower capacity utilisation rates in 1978.

The productivity movement can also be visualised in the figure. The 300 kt isoquant is the last one shown for 1966 in the top North-East part of the capacity region, while it is located almost in the middle (no. 8 from the start) for 1970, i.e., exhibiting a significant shift towards the origin.
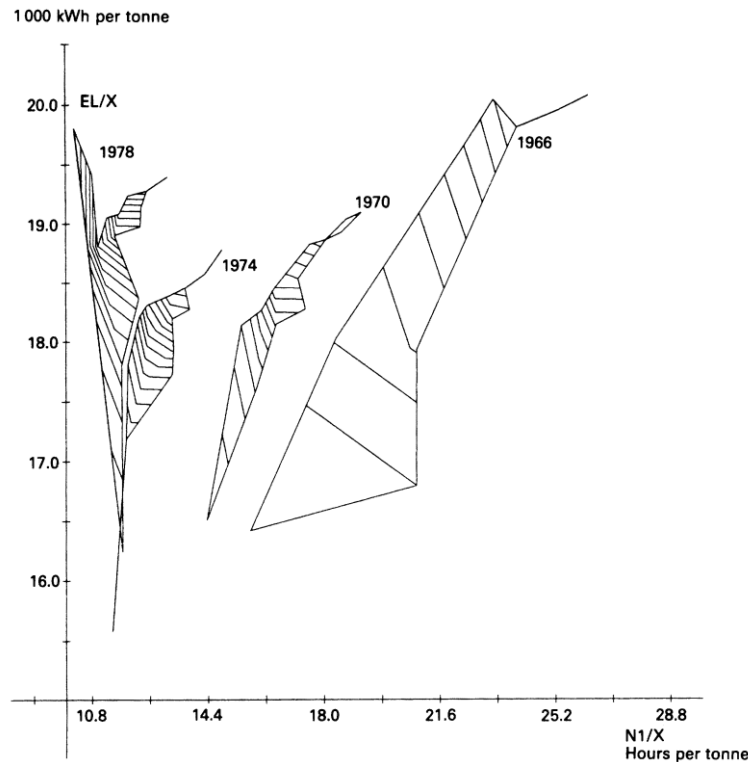
Source: Førsund and Jansen (1983)

Figure 4. The development of the capacity regions in input coefficient space
of the short-run industry function for Norwegian aluminium industry

The complete scheme

The putty-clay model may be too extreme to be empirically valid, at least on the firm level. In Johansen (1972) we find a general scheme for factors influencing technical change of the short-run industry function:

      a) Entry of new capacity reflecting the current ex ante function with embodied

         technical change and price expectations influencing factor ratios

      b) Exit of old capacity reflecting past choices of input coefficients

      c) Disembodied technical change

            i) Output increasing

            ii) Input saving

The new element is the possibility of disembodied technical change. The notion of disembodied technical change has partly to do with the level of aggregation. A firm may consist of several plants, and a plant may have several different pieces of equipment, reflecting the nature of the production taking place. When basic components within a firm or plant are substituted by new ones embodying technical improvements, we may at the more aggregated level call this technical change for disembodied. Disembodied technical change is assumed by many, and has been reported in empirical studies. How to distinguish between embodied and disembodied technical change is an empirical challenge (Belifante, 1978; Jorgenson, 1966; Fuss, 1978).

Førsund et al. (1996) tried to identify the elements of the complete scheme. Using the history of investment in the Finnish brewing industry for the period 1955 – 1984 three periods where identified; starting with a long period of average output growth and investments, followed by a short period of very rapid investment and output growth, coinciding with deregulation of the sector, and then followed by a period of quite modest growth in investment and output, coinciding with rapid change in relative price between labour and energy, constituting a period of consolidation after the investment boom. An ex ante function, with the possibility of separate parameter values for the first and the last periods and with the possibility of change in the elasticity of scale function and biased technical change, was estimated as a frontier function. The ex ante function was not attempted to be estimated for the period of the investment burst because of the possibility of a rapid embodied technical change in this short period. The results indicated that the optimal scale increased by 85% during the first period, while it jumped to the extent of 292% from the last year of the first period to the first year of the last period. During the latter consolidation period optimal scale remained almost constant.

The technical advance measured at optimal scale is 4.9% average unit cost reduction per year during the first period, 16% during the investment boom period, and below 1% during the consolidation period, and the latter increase due to the effect of biased technical change.

The productivity change of the short-run industry function is quite steady at a rate of 3.1% over the complete range of aggregate output levels. For the investment boom period the productivity changed at a rate of 10% for the most efficient part of aggregate output, and falling off for higher output levels, but still significantly higher than for the first period. The last period of investment consolidation shows a different pattern with 6% productivity growth for the most efficient part of output, and then actually increasing for higher output levels. An almost stationary ex ante technology goes together with a substantial productivity improvement of the productivity of the short-run industry function. This effect may be attributed to disembodied technical change, but may also be explained as efficiency improvement achieved when running in the new investments and improving the performance of a process with many links in a chain from raw materials to beer (Johansen, 1972; Eide, 1976).

5. The ex ante function and productivity measurement

Technical advance
In the putty-clay model it is change in the ex ante micro production function that constitutes new technological knowledge. It is therefore natural to have a measure of technical change based on the development of the ex ante function over time. Salter names this technical advance. This is not a productivity measure for a sector, but is the source of productivity growth within the sector when new units exhibiting the state-of-the-art technology enters the sector, or existing units are rebuild to adapt new technology.

A question is how to obtain detailed enough information enabling us to estimate statistically the ex ante function without utilising engineering knowledge directly. We must now have information also on the capital volume involved. One problem with existing approaches is the standard assumption when estimating a frontier function that each unit in a cross section may potentially be best practice. However, if embodiment of technology is sufficiently

predominant, then only units having the last vintage of capital should be considered. Using older units may create a bias. Older technology may become best practice technology if, e.g., factor ratios have changed and old vintages become frontier units by "default". Different factor ratios may be chosen using new technology because of biased technical change, and also due to relative price changes reducing the relative use of what is now the expected most expensive inputs in the future. The number of genuine units of the most recent vintage may be few, and have too equal factor ratios for a satisfactorily statistical estimation.

Another problem in the presence of embodied technical change is the measurement of capital. Book values or numbers calculated from using the perpetual inventory method may create a bias in measuring capital that it is difficult to predict the sign of, or do anything about. The increasing quality or productivity of new capital may not be reflected in the recorded values. The best strategy may be to insist on replacement values (Johansen and Sørsveen, 1967).

We may again apply Salter's measure of productivity based on the change in costs for fixed output levels and factor prices. One of the factors is now capital. In Førsund and Hjalmarsson (1987) it is proposed to measure the change in costs, not for fixed output levels, but for the optimal scale. The argument is that partial productivities are maximal at the optimal scale, i.e., the unit cost is minimised, and therefore is a natural benchmark.

The interpretation of the Malmquist productivity index

The Malmquist productivity index (Caves et al., 1982) has become popular within productivity studies in recent years. This productivity index is calculated with reference to a frontier function, usually with capital as one of the inputs. The use of a frontier function, valid for a micro unit, to calculate productivity for a whole sector consisting of N units, is in general somewhat problematic in view of the vintage model exposed above. Let us develop this theme in more detail.

The standard Malmquist productivity index, $M_i(1,2)$, for a unit, e.g., a firm, observed in two periods, 1 and 2, is defined as

$$M_i(1,2) = \frac{E_{i2}}{E_{i1}} \ , i = 1,...,N \tag{18}$$

where E_{i1} , E_{i2} are the Farrell technical efficiency measures, either input- or output-oriented, and where 1,2 appearing as arguments in the Malmquist index function, and 1 and 2 as

subscript of the efficiency measures, indicate output and input variables from the two periods. Following the original definitions in Farrell (1957) the efficiency measures are between 0 and 1, with 1 characterising efficient units. The index i represents the reference technology. This technology can be for period 1 or period 2, or technically speaking any technology that is regarded as the most suitable for the purpose at hand. But it is natural that the technology represents a frontier technology, because the Farrell efficiency measures are defined with a frontier technology as a reference:

$$E_{i1} = min_{\theta_i} \{\theta_i : y = f_i(\theta_i x_1,..,\theta_i x_n, \theta_i K)\},$$
$$E_{i2} = min_{\delta_i} \{\delta_i : (y/\delta_i) = f_i(x_1,..,x_n, K)\} \ (i = 1,..,N) \tag{19}$$

The output and the inputs are observed quantities for a micro unit. For simplicity we have used a single output production function $f_i(.)$ that is an ex ante micro production function of type (1) in Section 3. Generalisation to multiple outputs is straightforward. We assume that we have unique solutions for the scaling factors. No feasible point can have larger output given the inputs, or using less input given the output than points on the frontier.

The Farrell efficiency measures can be interpreted in a productivity context. For the input-oriented measure E_{i1} the proportional maximal scaling-back of the inputs may be regarded as expressing the productivity of the observation relative to the productivity of the benchmark point on the frontier function $f_i(.)$. Output is kept fixed, so the proportional scaling factor is the ratio between an input on the frontier and an observed input. Similarly, for the output-oriented measure E_{i2} observed inputs are fixed, so the scaling factor is the ratio of observed output and output on the frontier, i.e., a measure of relative productivity.

The family of Farrell efficiency measures is illustrated in Figure 5 (Førsund and Hjalmarsson, 1979) in the case of estimating the frontier within a non-parametric framework specifying a piecewise linear frontier function using linear programming. The approach is termed Data Envelopment Analysis (DEA). The point of departure is the observation P that is inefficient with respect to the two frontiers CRS (constant returns to scale) and VRS (variable returns to scale). The reference point on the frontier for the input-oriented measure E_1 with respect to the VRS frontier is P_1^{ref}, and P_1^{crs} with respect to the CRS frontier. The reference point on the frontier for the output-oriented measure E_2 with respect to the VRS frontier is P_2^{ref}, and P_2^{crs} with respect to the CRS frontier. 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 the reference
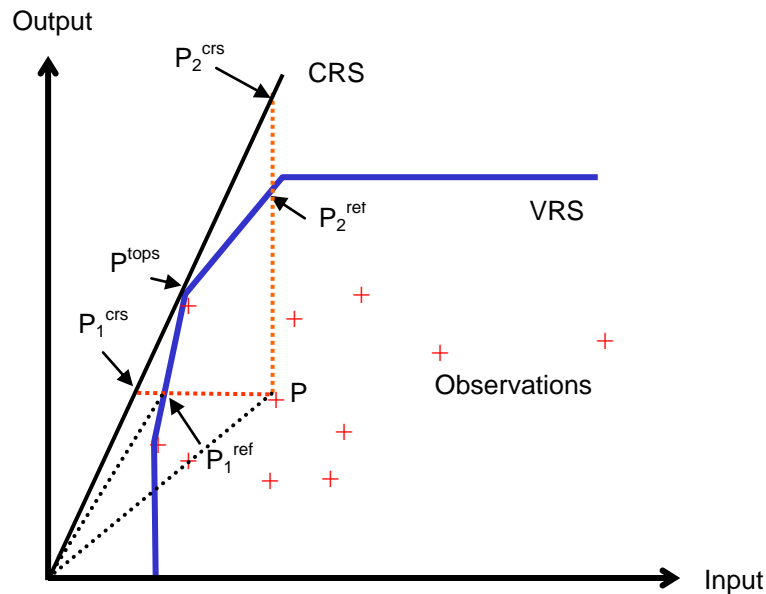
Figure 5. The family of Farrell efficiency measures

point on the VRS frontier gives the productivity of this reference point. In the case of one output and one input the interpretation of the Farrell efficiency measure as the relative productivity is exact. The same productivity interpretation holds for the output-oriented efficiency measures, and also using the CRS reference frontier. It is easy to see geometrically that in the case of using the CRS frontier the two efficiency measures must be identical, as pointed out by Farrell (1957).

As is easily seen from Figure 5 the productivity at the CRS frontier is maximal. Comparing the observation with the reference point P^{tops} therefore gives the relative productivity of an observation to the maximal productivity on the frontier. This efficiency measure E_3 is therefore termed the measure of technical productivity.[3] The measure is straightforwardly generalised to multiple outputs and inputs. The two remaining efficiency measures E_4 and E_5 are the (pure) scale efficiency measures comparing the productivity of the reference points P_1^{ref} and P_2^{ref} respectively with the point P^{tops} of maximal productivity on the frontier.

Returning to the definitions of the Malmquist index (18) we have that in the case of a CRS reference frontier and a single output and a single input the ratio of the efficiency measures

---

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

(either input- or output-oriented) is the ratio of productivity of an observation in period 2 relative to the productivity of the unit in period 1. However, in the general case of multiple outputs and inputs the productivity interpretation is more involved. Our conjecture is that a straightforward interpretation of productivity where the influence of the CRS reference frontier cancels out is only in the case of inverse homotheticity of the CRS frontier (Färe and Primont, 1995).

An illustration of the Malmquist index is provided in Figure 6. A unit is being observed in period 1 and in period 2, illustrated by P_1 and P_2. Let us assume that the relevant frontier technology is VRS. However, as choice of reference technology for productivity measurement (at least) two considerations have to be taken into account: the desired homogeneity property of the productivity index, and comparability of productivity changes between different periods.

As to the first consideration a TFP (total factor productivity) measure should be homogenous of degree 1; a doubling of outputs from one period to the next having the same level of inputs, should show up as a doubling of productivity. The VRS specification is then not suitable in general as a reference technology. It can be shown that the Malmquist productivity measure may be perversely influenced by scale if specifying VRS (Grifell-Tatjé and Lovell, 1995). But using the points of maximal productivity of the frontier technology will take care of this
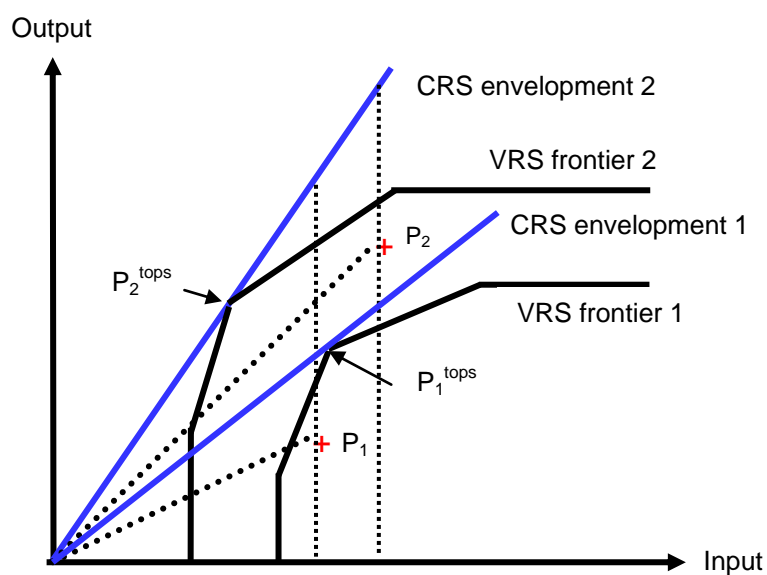

Figure 6. The Malmquist productivity index
obeying homogeneity of degree 1 and circularity

problem, and making the Malmquist productivity index homogeneous of degree 1. Therefore, by using the CRS envelopment as a reference technology, problems with impacts of scale changes are avoided and homogeneity of degree 1 is assured, since that maximal productivity is the characterisation of technically optimal scale, where the scale elasticity is equal to one and the production function is locally homogeneous of degree 1 (Frisch, 1965). However, using a CRS frontier as a reference does not mean that we assume CRS, it just serves as a reference for TFP measures. Then $E_1 = E_2 = E_3$, and it does not matter which orientation we use.

In order to compare productivity measures for different periods in a meaningful way, i.e., securing that the percentage changes refer to a common scale, the index must be circular (Gini, 1931). One way to achieve this is to use the same reference technology for all productivity measurements, i.e., the reference technology i in (18) is kept fixed when the successive periods 1 and 2 of comparison changes (Berg et al., 1992). The Malmquist index (18) can then be decomposed into efficiency change, or catching-up, and technology shift:

$$M_i(1,2) = \frac{E_{i2}}{E_{i1}} = \frac{E_{22}}{E_{11}} \frac{E_{i2} / E_{22}}{E_{i1} / E_{11}} \tag{20}$$

When a decomposition of the Malmquist index is performed into the components efficiency change (catching up) and technology shift (Färe et al., 1994), it is the latter that can be interpreted as the technical advance of the ex ante function.

Choosing the CRS envelopment for period 2 as the benchmark technology the Malmquist index can be identified in Figure 6 as the ratio of the efficiency score of observation P_2, using the benchmark technology (using the output orientation for convenience), and the efficiency score of observation P_1, using the benchmark technology. It should be easy to see that this measure is the same as forming the ratio of measuring the productivity of observation P_2 relative to the productivity of P_2^{tops}, and the productivity of observation P_1 relative to the productivity of P_2^{tops}. With only a single output and a single input the productivity at optimal scale cancels out. However, with multiple outputs and inputs the choice of benchmark technology will in general influence the values of the indexes, as mentioned above.

The first ratio in the last expression of (20) is the relative efficiency change when the efficiency scores are calculated relative to own-period technologies. The distance between the period technologies will in general express technical advance. In order to maintain circularity,

the last term in the last expression of (20), the technology-shift measure, has to be expressed as a double relativity; the efficiency scores calculated using the technology benchmark is expressed relative to the efficiency score calculated using the own-period technology as reference. The period reference technologies also have to be the CRS envelopes, as indicated in Figure 6. Using the period 2 CRS envelope as the reference technology, the figure (output orientation) shows the measure of technology shift as the relative distance between period 2 and period 1 CRS envelopes, measured through the observation for period 1.

Productivity for an industry is usually calculated as the, arithmetic or geometric, mean of the individual productivity changes. The main problem using the Malmquist index is that the interpretation of productivity change measured by the index (18) is not clear as to the production concept used. The frontier function is best used for measuring technical advance attached to single unit, and not for measuring productivity change for an industry. Use of the Malmquist index blurs the distinction between the ex ante micro function relevant for investments and the short-run production possibilities for the industry as a unit.

A standard assumption made when estimating the efficiency scores in the Malmquist index (18) is that any unit may potentially be a frontier unit. However, this is only the case if there are no vintage effects. This may be the case for pure service industries where capital equipment plays a minor role, but not in process industries, like pulp and paper, thermal electricity generation, cement, oil refineries, etc. where the Malmquist index has been applied in the literature.

Moreover, in the case of disembodied technical change that in principle can only be relevant for existing units, the use of the index cannot discriminate between efficiency change and disembodied technical change. This may be of concern for those that may miss the importance of efficiency improvements in the empirical literature on more aggregated productivity growth. A tentative distinction between efficiency improvements and disembodied technical change may be to say that efficiency improvements are managerial and do not entail any (at least serious) investments, while disembodied technical change require some, although may be modest, investments in order to be realised. Some small-step improvements here and there of a technical nature may sum up to measurable disembodied technical change in the short-run industry function. In Johansen (1972) it is pointed out that

such current improvements would be reasonable to assume being incorporated in the most recent blueprint of ex ante technology.

Calculating productivity using a Malmquist index based on a micro frontier production function construct, when vintage effects are important, cannot tell us correctly how the short-run industry-wide production possibilities develops regarding productivity. Using an ex ante micro frontier function as the crucial benchmark is not relevant for units embodying more or less outdated technology. Interpreting the Malmquist productivity index on an aggregate level is not consistent with the information provided by the process of dynamic structural change as set out in Figure 1.

The Malmquist index for a unit can at best be used as an indicator for how the productivity develops compared with a situation where the technology is continuously updated with the most modern capital. However, this is often very unrealistic when vintage effects are present (see Førsund and Hjalmarsson (1974) for an approach recognising vintages). The efficiency measure, or catching-up term, may serve as a description of structural change within the industry. If the term is greater (smaller) than one it means that the productivity of the observation has moved closer to (further away from) the own-period maximal frontier productivity. Whether the average measure over units has any useful policy interpretation is less clear.

6. Conclusions

We have reviewed a modelling framework for bottom-up analyses of structural change and productivity growth of an industry. Key elements in understanding structural change are the entering of capacity embodying new technology and exiting of capacity no longer able to yield positive profit. The point of departure for the vintage analysis of Johansen (1972) is the need for clarifying the production function concepts needed for understanding the structural change. Within his vintage approach he identified the ex ante micro unit production function as relevant when investing in new capacity, and the short-run industry production function giving the production possibilities at the industry level. In the extreme version of the vintage model there are neoclassical substitution possibilities between variable inputs and capital

before investing, but no substitution possibilities ex post. The micro unit production possibility ex post is described by a limitational law of fixed input coefficients. Moreover, capital is sunk cost, and has no role in the short-run industry production function, but the given output capacities decided upon at the time of investment is reflected in the structure of the production possibilities. Extending this model disembodied technical change is introduced in Johansen's complete scheme of all elements influencing the current structure of an industry.

Measuring productivity development within this framework means distinguishing sharply between technical advance measured by the change in the ex ante micro production function, and the productivity change measured for the short-run industry production function only concerned with variable inputs. Measurement of capital is usually a headache in productivity studies. The vintage approach does not conceptually need measures of capital. Concerning capital the focus will shift to measuring the return on capital investment. Data on quasi-rents should be used for such calculations.

A discussion of how to relate the increasingly popular use of the Malmquist productivity index to the vintage model uncovered some problems with the productivity interpretation of the Malmquist index when embodiment of technology is present. It was also pointed out that it is difficult to separate efficiency improvements that may be related to managerial performance, and disembodied technical change. These topics represent interesting research question to follow up.

The modelling framework discussed in the paper should be useful for policy purposes, especially in economies with a rapid growth in manufacturing exhibiting a vintage structure, like in China. A successful industrial policy cannot be implemented without understanding the dynamic forces at play, leading to significant reallocation of resources such as labour, materials and energy within an industry.

References

Ayres, R, U. (1999). The minimum complexity of endogenous growth models: the role of physical resource flows and technology, *Working Paper* 99/37/EPS, CMER, INSEAD, Fontainebleau.

Belifante, A. (1978). The identification of technical change in the electricity generating industry. In M. Fuss and D. McFadden (eds.): *Production economics: a dual approach to theory and applications*, Volume 2, Chapter IV.3, pp.149-186.

Berg, S. A., Førsund, F. R., Jansen, E. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980 – 1989. *Scandinavian Journal of Economics* 94, S211-S228, 1992.

Bye, T., Holmøy, E. and Massey K. H. (2006). Removing policy based comparative advantage for energy intensive production. Necessary adjustments of the real exchange rate and industry structure. Discussion Papers 462 - Statistics Norway, July 2006.

Caves, D. W., Christensen, L. R. and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50, 1393-1414.

Diewert, W. E. and Fox, K. J. (2008). On the estimation of returns to scale, technical progress and monopolistic markups. *Journal of Econometrics* 145, 174-193.

Eide, E. (1976). Do new factories embody *potential* best practice technology? *Empirical Economics* 1(2), 119-127.

Färe, R. and Primont, D. (1995). *Multi-output production and duality: theory and applications*. Boston/London/Dordrecht: Kluwer Academic Publishers.

Färe, R., Grosskopf, S. Lindgren, B. and Roos, P. (1994). Productivity developments in Swedish hospitals: a Malmquist output index approach. In Charnes, A., W. W. Cooper, A. Y. Lewin and L. M. Seiford (eds.): *Data Envelopment Analysis: Theory, Methodology, and Applications*, Boston/Dordrecht/London: Kluwer Academic Publishers, 253-272.

Farrell, M, J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *Series A (General)* 120 (III), 253-281.

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

Fuss, M. (1978). Factor substitution in electricity generation: a test of the putty-clay hypothesis. In Melvin Fuss and Daniel McFadden (eds.): *Production economics: a dual approach to theory and applications*, Volume 2, Chapter IV.4, pp.187-213.

Førsund, F.R. and Hjalmarsson, L. (1974). On the Measurement of Productive Efficiency. *Swedish Journal of Economics* 76 (2), 141-154.

Førsund, F.R. and Hjalmarsson, L. (1979). Generalised Farrell measures of Efficiency: An application to milk processing in Swedish dairy plants", *Economic Journal* 89, 294-315.

Førsund, F.R. and Hjalmarsson, L. (1987). *Analyses of industrial structure: a putty-clay approach*. Industrial Institute for Economic and Social Research, Stockholm: Almqvist and Wicksell International.

Førsund, F. R. and Jansen, E. S. (1983). Technical progress and structural change in the Norwegian aluminium industry. *Scandinavian Journal of Economics* 85, 113-126.

Førsund, F.R., Hjalmarsson, L. and Summa, T. (1996). The interplay between micro-frontier and sectoral short-run production functions. *Scandinavian Journal of Economics*, 98(3), 365-386.

Gini, C. (1931). On the circular test of index numbers. *Metron* (*International Review of Statistics*) 9(2), 3-24.

Grifell-Tatjé, E. and Lovell, C. A. K. (1995). A Note on the Malmquist Productivity Index. *Economics Letters* 47(2), 169-175.

Heckscher, E. (1918). *Svenska produktionsproblem* [Swedish production problems]. Stockholm: Bonniers.

Johansen, L. (1959). Substitution versus fixed production coefficients in the theory of economic growth. *Econometrica* 27(2), 157-176.

Johansen, L. (1972). *Production functions. An integration of micro and macro, short run and long run aspects*. Amsterdam-London: North-Holland.

Johansen, L. and Sørsveen, Å. (1967). Notes on the measurement of real capital in relation to economic planning models. *Review of Income and Wealth*, Series 13, 175-197.

Jorgenson, D. W. (1966). The embodiment hypothesis. *Journal of Political Economy* 74(1), 1-17.

Marshall, A. (1890). *Principles of economics,* 8th ed. (1966). London-Melbourne -Toronto: Macmillan.

Maywald, K. (1957). The best and the average in productivity studies and in long-term forecasting. The Productivity Measurement Review 9, 37-49.

Mitchell, (1937). The social sciences and national planning. In McKenzie, ed. *Planned society. Yesterday today tomorrow*. New York: Prentice Hall.

Phelps, E. S. (1963). Substitution, fixed proportions, growth and distribution. *International Economic Review* 4, 265-288.

Salter, W.E.G. (1960). *Productivity and technical change*. London: Cambridge University press.

Wedervang F. (1964). *Development of a population of industrial firms*. Bergen: Scandinavian University Books.

Wohlin, L. (1970). *Skogsindustriens strukturomvandling och expansionsmöjligheter*. [Forest-based industries: structural change and growth potentials, in Swedish with English summary]. Stockholm: Industriens utredningsinstitut.

Åkerman, G. (1931). *Den industrielle rationaliseringen och dess verkningar* [The industrial rationalisation and its effects]. SOU 32, Stockholm.