Industriens Utredningsinstitut
THE INDUSTRIAL INSTITUTE FOR ECONOMIC AND SOCIAL RESEARCH

A list of Working Papers on the last pages

No. 300, 1991

AN INDIRECT APPROACH TO MEASURING PRODUCTIVITY IN PRIVATE SERVICES

by

Erik Mellander

October 1991
Revised, March 1992

2.92 10.000

Postadress
Box 5501
114 85 Stockholm

Gatuadress
Industrihuset
Storgatan 19

Telefon
08-783 80 00
Telefax
08-661 79 69

Bankgiro
446-9995

Postgiro
19 15 92-5

An Indirect Approach to Measuring Productivity in Private Services

by

Erik Mellander*

The Industrial Institute for Economic and Social Research (IUI)

Box 5501, S – 114 85 Stockholm

Fax: + 46 – 8 – 661 79 69

March 1992

Abstract: Productivity measurement is considered in the context of incomplete output information. Only a *value* measure of output is assumed to be available, which is typical for many service industries. Input markets are assumed competitive while the output market is allowed to be non-competitive, potential markups being assumed to be either known or constant. It is shown that if production technologies are homothetic and the elasticities of total costs with respect to output are strictly increasing, the given data are equivalent to complete information, provided the markup is known. If it is not the results hold *conditional* on the unknown markup.

* Financial support from the *Bank of Sweden Tercentenary Foundation*, the *Royal Swedish Academy of Sciences*, and the *Sweden — America Foundation* is gratefully acknowledged. I have benefitted from comments from Thomas Andersson, Ernst Berndt, Rolf Färe, Lars Grönstedt, Sten Nyberg and, in particular, Fabio Schiantarelli and two anonymous referees.

I. Introduction

A salient feature of the service industry is that, in general, it is very difficult to measure its output. While input data mostly are available, it is commonly the case that reliable output quantity or output price data cannot be found in official statistical sources. Indeed, for some types of services, notably within the public sector, it is not even possible to obtain output value measures. The purpose of an earlier paper [Mellander and Ysander (1990)] was to examine what conclusions that can be drawn about the production technology and the producer behavior in the latter situation, i.e. when there is no output information whatsoever. It was shown that for homothetic production technologies – i.e. technologies which have the property that the optimal factor mix is independent of the level of production – time series on input prices and input quantities only can be used to study almost all dimensions of the production process by means of a dual approach, using the cost function as the instrument of analysis. In principle, the only aspects that cannot be investigated are those affecting input demands neutrally, e.g. (purely) Hicks-neutral technical change and properties relating to returns to scale.[1]

This paper extends the analysis to the case when a measure of the value of output is available, which is the typical situation for most kinds of private services. Thus, the analysis will again be based on the cost function and it will presume that firms are endowed with homothetic production technologies and operate on a competitive input market. The output market, on the other hand, will be allowed, to be non – competitive. It will be assumed, however, that in the context of markup pricing the markup is either known or (approximately) constant. Thus, while there can be a wedge between the marginal cost and the output price it is assumed that if this wedge is unknown it can be treated parametrically. Finally, to

[1] As a reminder, for a homothetic technology returns to scale are determined solely by the level of output and so will be independent of the input mix.

simplify the analysis, the producer is assumed to maximize profits and attention is confined to static equilibrium models.

The homotheticity assumption and the requirement that, in the absence of *a priori* information, the possible markup be constant might perhaps seem rather restrictive. It should be noticed, however, that if one instead follows the traditional route in dealing with the output measurement problem and replaces the unknown output by some proxy variable(s) then it is usually impossible to say anything about under which circumstances the variations in the proxy (-ies) really mirror the changes in the actual output, and so one can never be sure whether the results obtained are valid. Here, it is completely clear under which conditions the method suggested is applicable.

Concerning the homotheticity assumption it can be argued that it is more easily justified in the context of service production than in goods production. Due to the more limited scope for automatization in the service industry, expansion often takes place by the setting up of additional production units (offices), similar to the ones already existing. Examples can be found within the banking industry and in travel agencies, for instance. As a result, the input proportions change much less than when the expansion occurs mainly through additions to the capital stock, as is the case in the manufacturing industry.[2] Regarding the constant markup assumption it should be noticed that in the context of productivity measurement it is quite common to assume not only the input but also the output market to be competitive. Here, the latter assumption is relaxed, albeit in a crude way.

The paper unfolds as follows. Section II starts with a description of the model in terms of the firms' production technologies and the market conditions. The existence of an equilibrium in the output market is then established and sufficient

---

[2] Of course, input proportions change over time in the service industry, too, because of changes in relative factor prices. The claim here is simply that the smaller changes observed for the service industry as compared to the manufacturing industry are due to the fact that *ceteris paribus* expansion affects input proportions much less in the service industry than in the manufacturing industry.

conditions for a unique and stable equilibrium are considered. Finally, the key result of the paper is derived, namely a relation between the unknown output variable and the ratio of the value of output over total costs, for which data are assumed to be available. In Section III a decomposition of total factor productivity growth into the effects of (Hicks-)neutral technical change, biased technical change, and effects from returns to scale is considered, followed by a general discussion of the estimation of each of these components. Section IV demonstrates how the theoretical results can be implemented by means of particular flexible functional form, namely the translog cost function. Concluding comments are given in Section V.

II. The model

The basic structure of the model is given by the following three sets of assumptions.

(i) *Technological assumptions*

Firms, indexed by $i = 1,...,m$ where m might be equal to 1, are assumed produce the (homogeneous) output good by means of (possibly different) homothetic technologies. Given cost minimization [cf. (iii) below] the firm's technology can be characterized by means of the cost function which. Due to the homotheticity assumption, the cost function is separable in output, y_i, and the vector $\boldsymbol{w}$ of input prices, according to

$$C_i = C_i(y_i, \boldsymbol{w}, t) = f_i(y_i) \cdot C_i(1, \boldsymbol{w}, t)\ , \quad i = 1,...,m, \tag{1}$$

where the time index t represents the state of the technology and $C_i(1, \boldsymbol{w}, t)$ denotes the cost of producing one unit of output. Concerning notation, C_i will only be used to denote total costs, to avoid confusion between total and unit cost. Further, boldface types will be used to denote vectors (small letters) and matrices (capitals). Accordingly, it is assumed here that output can be treated as a scalar. This does not exclude multiple output activities but it requires the existence of an output aggregate.[3]

The function $f_i(y)$, which is monotonically increasing, determines the scaling properties of the technology.[4] In this and the following section C_i will be assumed to be a regular cost function.[5] In addition, it will be taken to be twice differentiable with respect to each of its arguments.

It is further assumed that the elasticity of total costs with respect to output

$$\varepsilon_i = \varepsilon_i(y_i) \equiv \frac{\partial \ell n C_i(y_i, w, t)}{\partial \ell n y_i} = \frac{y_i \cdot f_i'(y_i)}{f_i(y_i)}, \quad i = 1,...,m, \tag{2}$$

is monotonically increasing in output, i.e.,

$$\varepsilon_i'(y_i) > 0 \quad \forall\, y_i\,, \quad i = 1,...,m. \tag{3}$$

While the cost/output elasticity very often is assumed to be non-decreasing, it is less often assumed to be strictly increasing since this rules out homogeneous technologies and hence, in particular, technologies that are homogeneous of degree 1, i.e. exhibit constant returns to scale.[6] The reason for the strict monotonicity assumption is that in the discussion below the existence of a mapping from the cost/output elasticity to the level of output will be exploited; such a mapping exists if, and only if, the function $\varepsilon_i(y_i)$ can be inverted, i.e. if it is strictly monotonic.[7]

---

[3] This, in turn, amounts to assuming that the optimal output proportions (but not the levels) can be determined without any input information.

[4] In principle, it is conceivable that technological developments might affect the technology's scaling properties, in which case t should be an argument also in the f_i function. For simplicity, I abstract from that possibility here.

[5] Regularity conditions can be found, e.g., in Diewert (1971). Some of these conditions can be tested statistically, cf. Section IV.

[6] In principle, the only troublesome fact is that constant returns to scale technologies cannot be considered. As technologies that are homogeneous of degree $r \neq 1$ have ever-increasing or ever-decreasing returns to scale and, hence, lack well defined optimal levels of production, the exclusion of them is not very serious.

[7] This is not to say that homogeneous technologies cannot be analyzed at all – the results in Mellander and Ysander (1990) are valid for homogeneous technologies,

Finally, it is assumed that marginal costs are strictly increasing, i.e.

$$f_i''(y_i) > 0 \ \forall \ y_i \ , \quad \mathrm{i} = 1,...,\mathrm{m}. \tag{4}$$

(ii) *Assumptions about market conditions*

Input markets are assumed to be competitive while the output market is allowed to be non-competitive. The inverse industry market demand curve $p(\Sigma_{i=1}^{n} y_i) = p(1^t y)$, where superindex "t" denotes transpose, is assumed to be finite valued, non-negative, strictly decreasing, and twice differentiable.[8] Moreover, total industry revenue, i.e. $1^t y \cdot p(1^t y)$, is assumed to be bounded and strictly concave for all y.

(iii) *Assumptions about information sets and behavior*

All firms are assumed to know the inverse industry market demand curve, their own cost function, and the cost functions of all other firms. Given this information they seek to maximize profits. □

The assumption that production technologies are homothetic implies that the profit maximization problem of firm i can be divided into two separate subproblems. The first problem is to choose the cost-minimizing factor proportions, which are independent of the scale of production. The second problem is to choose the optimal level of output.[9] The solution to the first problem is given by $C_i(1,w,t)$. When solving the second problem the firm can take $C_i(1,w,t)$ as given. Accordingly, firm i's maximization problem can be written

too. It means, however, that for these technologies the output value measure yields no extra information in addition to that provided by the input data.

[8] Of course, the argument list of the price function will in general include a number of exogenous shift variables. To simplify the notation, these are suppressed here.

[9] For a non-homothetic technology it is not possible to separate these two problems as the cost-minimizing factor mix will be dependent upon the level of production. The fact that I denote the problems the "first" and the "second", respectively, should not be taken to indicate anything about the order in which they are to be solved; as will be seen below it is perfectly possible to begin by considering the second one.

$$\max_{y_i} \ \pi_i = p(1^t\boldsymbol{y}) \cdot y_i - f_i(y_i) \cdot C_i(1,\boldsymbol{w},t) \ .$$

It should be noticed assumptions (i) and (ii) imply that the profit function π_i is strictly concave with respect to y_i .[10]

Following Appelbaum (1982), the first order conditions for profit maximization can be formulated according to

$$p(1 - \theta_i \eta) = f_i'(y_i) \cdot C_i(1,\boldsymbol{w},t) \ , \quad i = 1,\ldots,m, \tag{5}$$

where

$$\theta_i \equiv (\partial 1^t\boldsymbol{y}/\partial y_i) \cdot (y_i/1^t\boldsymbol{y}) \tag{6}$$

is the conjectural elasticity of total industry output with respect to the output of firm i and η is the inverse demand elasticity, defined as

$$\eta = -[\partial p(1^t\boldsymbol{y})/\partial 1^t\boldsymbol{y}] \cdot [1^t\boldsymbol{y}/p(1^t\boldsymbol{y})] \tag{7}$$

According to (6), the firm should set its output such that its marginal cost equal its *perceived* marginal revenue. This formulation of the first order condition is consistent with a wide range of behavioral modes. E.g., under Cournot behavior the conjectural variation $(\partial 1^t\boldsymbol{y}/\partial y_i)$ equals one implying that the conjectural elasticity θ_i reduces to the output share of firm i. In the case of perfect competition $\theta_i = 0$ for all i. Further, under pure monopoly and in the case of collusive behavior the conjectural elasticity will be identically equal to one, in the former case because $y_1 = 1^t\boldsymbol{y}$ and in the latter because $(\partial 1^t\boldsymbol{y}/\partial y_i) = 1^t\boldsymbol{y}/y_i$ for all i. Although other types of behavior are also conceivable within this framework only the four types just mentioned will be considered here.

Concerning the Cournot oligopoly game the analysis in Friedman (1986, pp.

[10]The strict concavity of total industry revenue with respect to total output implies that $p(1^t\boldsymbol{y}) \cdot y_i$ is strictly concave with respect to y_i . Further, (4) implies that the cost function is strictly convex with respect to y_i or, equivalently that the negative of the cost function is strictly concave with respect to y_i . Accordingly, π_i is a sum of two strictly concave functions and so must itself be strictly concave.

54 – 56) demonstrates the existence of at least one equilibrium point.[11] Since pure competition can be viewed as a limiting case of the Cournot oligopoly it follows that there must be at least one pure competition equilibrium, too. In the context of pure monopoly the existence of equilibrium is trivial. To the extent that the case of collusive behavior can be treated like a multi – plant monopoly operation, i.e. if agreements can really be considered binding, it is clear that there must exist an equilibrium in that case, too.

In the Appendix conditions for the equilibrium to be unique and stable are considered for the the simple case when the inverse demand curve is linear. (As is well known, Cournot behavior and collusion yield the same outcome in this case.) It is shown that under this assumption the equilibrium is unique if there are two firms. For $m = 3$ it is demonstrated that, essentially, the equilibrium is unique if the slope of the demand curve is less than twice the geometric mean of the slopes of the firm's marginal cost curves in absolute value. For stability it is required, in addition, that the slopes of the firms marginal cost curves exceed the absolute value of slope of the demand curve if $m = 2$. If $m = 3$ the slopes of the marginal cost curves have to be at least twice the absolute value of the slope of the demand curve.

An equilibrium relation will now be derived between the output level y_i – which is presumed to be unknown to the econometrician – and total costs C_i and the value of output $V_i \equiv p \cdot y_i$, for which data are assumed to be available. The first step is to solve (5) for p, yielding

$$p = \kappa_i \cdot f_i'(y_i) \cdot C_i(1, w, t) \ , \quad i = 1,\ldots,m \qquad (8)$$

[11] Taken together, assumptions (i) – (iii) fulfill Conditions 2.1 – 2.3 in Friedman, with one minor qualification: whereas in Friedman both the inverse demand function and the cost functions are defined over the range $[0, \infty)$ the corresponding range is here assumed to be $[\delta, \infty)$ where δ is some (infinitely) small positive number. The reason for this difference is that the dual cost function is well defined only for strictly positive output levels, cf. Diewert (1971, p. 489). It should also be said that since the intention is to use this model for measuring productivity developments one has to think of the Cournot one – shot game as being repeated over time.

where

$$\kappa_i \equiv \frac{1}{1 - \theta_i \eta}.$$

According to (8), the output price is given by a markup κ_i over marginal cost. It can easily be shown that $\theta_i \eta$ belongs to the the half-open interval $[0,1[$; cf. Appelbaum (op. cit. p. 290). Hence, κ_i will be bounded from below by 1, which is its value under perfect competition (since under perfect competition $\theta_i = 0$ for $i = 1,\ldots,m$). The markup will be highest in the contexts of pure monopoly or collusive behavior since in these cases the markup will be equal to $(1 - \eta)^{-1}$. In the Cournot case, finally, the markup will lie between these two extremes.

It will be assumed that if κ_i is unknown it can be treated as a constant. Note that, in general, this is *not* the same thing as assuming the price elasticity of demand to be constant; constancy of η is neither a necessary nor a sufficient condition for constancy of κ_i. However, if η is constant then, for κ_i to be constant, θ_i must be constant, too.[12]

The definition of V_i and (8) imply

$$V_i = \kappa_i \cdot f_i'(y_i) \cdot C_i(1,w,t) \cdot y_i \,, \quad i = 1,\ldots,m. \tag{9}$$

Hence, by (1) and (2)

$$\frac{V_i}{C_i} = \kappa_i \cdot \varepsilon_i (y_i) \,. \tag{10}$$

Since, according to (3), the function $\varepsilon_i (y_i)$ is invertible (10) implies that it is possible to express y_i in terms of the ratio V_i/C_i and the markup factor κ_i.

[12] The assumption of a constant κ_i may not be a too bad approximation even if η changes over time because such changes are likely to be counteracted be changes in θ_i of the opposite sign. E.g., take $p(\mathbf{1}^t\boldsymbol{y})$ to be linear. If at a given demand new firms enter – e.g. because costs have been reduced by technical change – then the new equilibrium will be characterized by a higher η than the old one but also by lower θ_i's, at least given Cournot behavior. Conversely, if at a given industry supply the demand curve shifts outward then new firms are likely to enter and the new equilibrium will have the same qualitative properties as in the first case.

This is the key result of the paper; the next two sections will discuss how it can be used in the estimation of total factor productivity growth.[13]

III. On the estimation of total factor productivity growth

In this and the following section the data available to the econometrician will be assumed to refer either to a single firm or to an aggregate of firms.[14] Accordingly, the firm index i will be dropped in the following.

The following (time series) information is assumed to exist. All relevant input data are known, i.e. both the quantities used of the n factors of production, $\boldsymbol{x} = (x_1,...,x_n)$, and the corresponding input prices $\boldsymbol{w} = (w_1,...,w_n)$, and, consequently, total costs $C \equiv \boldsymbol{w}'\boldsymbol{x}$. Regarding the output side, only the value V $(\equiv p \cdot y)$ of output is assumed to be known.

The following duality result, due to Ohta (1975), provides a useful decomposition of the growth in total factor productivity *(TFP)*. Denote the production function corresponding to $C(y,\boldsymbol{w},t)$ by $\psi(\boldsymbol{x},t)$. The rate of change in *TFP* can then be written

$$\hat{TFP} \equiv \frac{\partial \ln \psi(\boldsymbol{x},t)}{\partial t} = \nu \cdot \varepsilon^{-1}, \tag{11}$$

where

$$\nu \equiv -\frac{\partial \ln C(y,\boldsymbol{w},t)}{\partial t} \tag{12}$$

and ε^{-1} is the inverse of the elasticity of total costs with respect to output, defined in (2). The factor ν is the dual rate of technical change. Thus, if technical change

---

[13] I was surprised to find that the interesting relation (9) seems to have gone almost unnoticed. However, in a different context Morrison (1992, p. 55), considers the corresponding result in the monopoly case, i.e. when $\kappa = (1-\eta)^{-1}$.

[14] For the latter case to be meaningful the existence of a representative firm has to be assumed. As a discussion of aggregation conditions is outside the scope of this paper, suffice it to note that the assumption that all the m firms are identical is (trivially) sufficient for the existence of a representative firm.

has a positive impact ν measures the resulting rate of diminution in total costs. The inverse of the cost/output elasticity is the dual rate of return to scale. Returns to scale are increasing if $\varepsilon^{-1} > 1$, constant if $\varepsilon^{-1} = 1$, and decreasing if $\varepsilon^{-1} < 1$.

The dual rate of technical change can be further decomposed into two components corresponding to (Hicks-)neutral technical change and non – neutral, i.e. input specific, technical change. The former is a function of t while the latter depends on both t and w. Thus, denoting these functions by g and h,

$$\nu = -\frac{d\ln g(t)}{dt} - \frac{\partial \ln h(w,t)}{\partial t}. \tag{12'}$$

The problem of estimating the three components in *TFP* growth will now be examined in some detail. The estimation of the dual rate of return to scale, which does not require any assumptions about the functional form of the cost function, is discussed first. Concerning the two components relating to technical change, the one corresponding to non-neutral, i.e. input specific, technical changes will be very briefly considered, as its estimation is discussed in Mellander and Ysander (op. cit.). Finally, the direct relation between the presumed output measurement problem and the estimation of (Hicks-)neutral technical change will examined.

The dual rate of return to scale

Since the dual rate of return to scale is simply the inverse of the cost/output elasticity (10) yields

$$\varepsilon^{-1} = \kappa^{-1} \cdot \frac{C}{V}, \tag{13}$$

showing that the cost value ratio is proportional to the dual rate of return to scale. Accordingly, if the markup κ is known the dual rate of return to scale can be computed directly by means of the given data on total costs and the value of output.

However, if κ is not known *a priori* it is clear that the dual rate of returns to scale effect on total factor productivity can only be measured *conditional* upon this unknown constant. That is to say, it will be necessary to perform some kind of sensitivity analysis where the consequences of different assumptions about the magnitude of κ are investigated.

The dual rate of non–neutral technical change

Estimation of the last term in (12'), i.e. $-\partial \ell n h(w,t)/\partial t$, merely requires input data and the specification of an explicit functional form for the function $h(w,t)$. According to *Shephard's lemma*

$$S_j = \frac{\partial \ell n C}{\partial \ell n w_j} = \frac{\partial \ell n h(w,t)}{\partial \ell n w_j}, \qquad j = 1,\dots,n\,, \tag{14}$$

where S_j is the cost share of input j, i.e. $S_j \equiv (w_j x_j / C)$. Thus, the homotheticity assumption makes the cost shares functions of w and t only.

Imposing linear homogeneity of $h(w,t)$ in w, one can obtain an estimate of $h(w,t)$ by simultaneous estimation of n-1 of the n share equations.[15] Partial differentiation of this estimate with respect to t then yields an estimate of $\partial \ell n h(w,t)/\partial t$.[16]

Before turning to the estimation of the dual rate of Hicks – neutral technical change it should be said that while the above discussion has shown that the *minimal* requirements for the estimation of $-\partial \ell n h(w,t)/\partial t$ are very limited, the efficiency of the parameter estimates might be substantially increased if the cost function is estimated together with the system of cost shares.[17] Hence, the necessity to specify

---

[15] As is well known, the system of cost shares is singular and so one of the cost shares has to be left out in the estimation.

[16] Estimation of the system of input cost shares will also yield estimates of the Binswanger (1974) measures of the bias in technical change, and of elasticities of substitution and price elasticities; see Mellander and Ysander (1990) for a further discussion.

[17] Compared to estimation only of the system of cost shares, simultaneous estmation of the cost function and the cost shares will increase the efficiency of all the paramater estimates – and hence, in particular, those associated with the function $h(w,t)$ – for two reasons. First, parametrical constraints between the cost and the share equations will be taken into account explicitly in the latter case. Secondly, the residual in the cost function and the residuals in the share equations are probably correlated and this can be taken into account, too.

the cost function completely to enable estimation of the Hicks – neutral component in *TFP* growth, to be discussed next, has the positive side – effect of increasing the precision in the estimate of the function $h(w, t)$.[18]

The dual rate of neutral technical change

As the input cost shares are unaffected by neutral technical change estimation of the function $g(t)$ requires specification and estimation of the complete cost function. But estimation of the cost function presupposes data on y – or at least data providing information about the variation in y. This is the reason for assumption (3) which, through (9), ascertains that y can be expressed in terms of V, C, and κ.

One possibility is to assume that ε is linear in y, i.e.

$$\varepsilon = \beta + \varphi \cdot y, \quad \varphi > 0, \tag{15}$$ [19]

where the positivity constraint follows from (3). By (10),

$$y = -\frac{\beta}{\varphi} + \frac{1}{\kappa \cdot \varphi} \frac{V}{C} .$$

The specification (15) thus results in y becoming an affine transformation of the V/C – ratio. One can go one step further, however, by exploiting the fact that for empirical implementations it is the *variation* in y (rather than its level) that is of interest. The reason is that the explicit cost functions used in empirical applications constitute first or second order approximations to the "true" cost function around some point of expansion. Accordingly, what matters are the variations around the expansion point, which means that y (as well as the w_j's and t) are appropriately measured in terms of deviations from this point. The

[18] Moreover, consideration of the whole cost function in the estimation also makes it possible to test the validity of the restriction that $h(w, t)$ be linearly homogeneous in w. This is not possible if only the system of cost shares is estimated, since in that case the homogeneity restriction has to be imposed *a priori* (cf. above) to ascertain that the parameters to be estimated are identified..

[19] This type of specification has been discussed by Zellner and Revankar (1969). The corresponding scaling function is given by $f(y) = y^{\beta} \cdot \exp(\varphi \cdot y)$.

specification (15) may thus be reparameterized according to

$$\varepsilon = \lambda + \varphi\cdot(y - y_o)\ , \quad \varphi > 0\ , \tag{15'}$$

where $\lambda \equiv \beta + \varphi\cdot y_o$ and y_o denotes the point of expansion.

Since the choice of expansion point is arbitrary one can simply choose the one most convenient to work with. In the following, y_o will be thought of as being equal to the value on y in some "base-year", e.g. the mid-point of the observation period. However, y_o might equally well be set equal to the observation period mean, for instance.

By evaluating (15') at y and y_o, applying (9) twice, and forming the difference between the results one obtains

$$y - y_o = \frac{1}{\kappa\cdot\varphi}[(V/C) - (V/C)_o]\ , \quad \kappa \geq 1,\quad \varphi > 0\ , \tag{16}$$

where $(V/C)_o$ denotes the value/cost – ratio corresponding to the expansion point, i.e. its base-year value. Thus, by confining the attention to the deviation of y from the expansion point one arrives at a *proportional* relationship between $[(V/C) - (V/C)_o]$, for which data are assumed to be available, and the unknown output variable.

Of course, there are other specifications of ε which also have the property that ε is monotonically increasing in y. E.g., in studies based on the translog function proposed by Christensen *et al.* (1973) the following formulation is the most common one

$$\varepsilon = \alpha + \gamma\cdot(\ln y - \ln y_o) = \alpha + \gamma\cdot\ln(y/y_o) \quad \gamma > 0\ . \tag{17}$$

Subjecting (17) to the same operations as those performed on (15') to obtain (16) one gets

$$\ln(y/y_o) = \frac{1}{\kappa\cdot\gamma}[(V/C) - (V/C)_o]\ , \quad \kappa \geq 1,\quad \gamma > 0\ . \tag{18}$$

This is the specification that will be used in the next section. As a matter of matter of interpretation, note that by taking y_o to be the value of y in a base-year the left hand side of (18) becomes (the logarithm of) a quantity index for output.

IV. Implementation by means of the translog cost function

Since it is desirable to impose few *a priori* restrictions on the substitution possibilities among the factors of production, one should preferably consider flexible functional forms in the specification of an explicit functional form for the general cost function (1). The reason why the translog has been chosen here is that it is convenient to work and has been shown to provide adequate estimates of quite complex technologies [cf. Guilkey and Lovell (1980)]. It should be stressed, however, that, in principle, the above results can be implemented by means of any cost function which fulfills the assumptions in Section II.

The translog cost function constitutes a second order approximation to $\ell nC(y,w,t)$ in terms of $\ell n y$, $\ell n w_1,\dots,\ell n w_n$, and t. Denote the point around which the "true" cost function is expanded by $(\ell n y_o, \ell n w_{1o},\dots,\ell n w_{no}, t_o)$. The homothetic translog cost function can then be written

$$\ell nC = \ell nC(y,w,t) = \alpha_o + \ell n f(y) + \ell n g(t) + \ell n h(w,t) \tag{19}$$

where

$$\ell n f(y) \equiv \alpha_y \cdot \ell n \underline{y} + \tfrac{1}{2} \cdot \gamma_{yy} \cdot (\ell n \underline{y})^2 \tag{20}$$

$$\ell n g(t) \equiv \alpha_t \cdot \underline{t} + \tfrac{1}{2} \cdot \gamma_{tt} \cdot \underline{t}^2 , \tag{21}$$

$$\ell n h(w,t) \equiv \sum_{i=1}^{n} \alpha_i \cdot \ell n \underline{w}_i + \tfrac{1}{2} \cdot \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \cdot \ell n \underline{w}_i \ell n \underline{w}_j + \sum_{i=1}^{n} \gamma_{it} \cdot \underline{t} \cdot \ell n \underline{w}_i \right] . \tag{22}$$

and, for notational brevity,

$$\underline{y} \equiv y/y_o\ , \tag{23a}$$

$$\underline{w} \equiv (\underline{w}_1\ ,...,\underline{w}_n) \equiv (w_1/w_{1o}\ ,...,\ w_n/w_{no})\ , \tag{23b}$$

$$\underline{t} \equiv t - t_o\ . \tag{23c}$$

Thus, output and the input prices are taken to be measured on index form y_o and $w_{1o},...,w_{no}$ being the base year values, i.e. the values at time t_o .

Direct application of the last result in the previous section yields

$$\ell n \underline{y} = \kappa^{-1}\gamma_{yy}^{-1}\cdot \underline{q}\ , \tag{24}$$

where

$$\underline{q} \equiv (V/C) - (V/C)_o\ . \tag{25}$$

By means of (24) the cost function can be formulated in terms of V/C, $\underline{w}$, and $\underline{t}$, rather than y , w, and t , according to

$$\ell n C(V/C,\underline{w},\underline{t}) = \alpha + \ell n f^{\star}(V/C) + \ell n g(\underline{t}) + \ell n h(\underline{w},\underline{t})\ , \tag{26}$$

where

$$\ell n f^{\star}(V/C) = \alpha_q \cdot \underline{q} + \tfrac{1}{2}\cdot \gamma_{qq}\cdot \underline{q}^2 \tag{27}$$

and

$$\alpha_q = \alpha_y/(\kappa\cdot \gamma_{yy})\ , \tag{28}$$

$$\gamma_{qq} = 1/(\kappa^2\cdot \gamma_{yy})\ .^{20} \tag{29}$$

From (28) and (29) it is clear that for a given κ both α_y and γ_{yy} are identified.

To ensure that the cost/output elasticity implied by the model, i.e.

$$\varepsilon = \alpha_y + \gamma_{yy}\cdot \ell n \underline{y}\ , \tag{30}$$

[20] If the markup is known, the argument $\underline{q}/\kappa$ is substituted for $\underline{q}$ in the cost function, the parameters α_q and γ_{qq} becoming a_y/γ_{yy} and $1/\gamma_{yy}$, respectively.

is consistent with (9) and that the property (3) holds the following constraints must be imposed on α_q and γ_{qq}

$$0 < \alpha_q . \tag{31}$$

$$\gamma_{qq} = \alpha_q \cdot [(V/C)_o]^{-1} , \tag{32}$$

Together, (28), (29), (31), and (32) yield

$$\alpha_y = \kappa^{-1} \cdot (V/C)_o \quad (> 0), \tag{33}$$

$$\gamma_{yy} = \kappa^{-2} \cdot \alpha_q^{-1} \cdot (V/C)_o \quad (> 0). \tag{34}$$

By inserting the expressions for $\ell n \underline{y}$, α_y , and γ_{yy} given by (24), (33) and (34) in (30) one can easily verify that the equality $\varepsilon = \kappa^{-1} \cdot (V/C)$ will always hold, as required by (10). Moreover, (3) holds by the positivity of γ_{yy} . Of course, that (3) and (10) hold does not mean that the empirical implementation of the model does not yield any new information about the technology's scaling properties; it will result in estimates of the scaling parameters α_y and γ_{yy} (conditional on κ) and, moreover, it will make it possible to form some idea about the precision in the estimate of the scale elasticity, through the standard error of the parameter α_q of which γ_{yy} is a function. Finally, it should be remembered that it is only by using $\underline{q}$ as an instrument for $\ell n \underline{y}$ that one can estimate the function $g(t)$.

Unfortunately, it is not possible to impose *a priori* constraints on the parameters such that (4) is guaranteed to hold. It can be concluded, however, that (4) implies an upper bound on α_q which should be approximately $4 \cdot (V/C)_o$.[21]

In empirical applications, the cost function is estimated jointly with n-1 of the input cost shares, given by [cf. (14)]

[21]This conclusion is obtained as follows. By direct calculation it can be shown that (4) holds if and only if $(\varepsilon^2 - \varepsilon + \gamma_{yy}) > 0$ implying that $\gamma_{yy} > 0.25$ is a *sufficient* condition. By inspection of (34) it can be seen that $\gamma_{yy} > 0.25$ translates into the condition $\alpha_q < 4 \cdot (V/C)_o \cdot \kappa^{-2}$. Since $\kappa \geq 1$, $\alpha_q < 4 \cdot (V/C)_o$ is necessary for $\gamma_{yy} > 0.25$ which in turn is sufficient for (4).

$$S_j = \alpha_j + \tfrac{1}{2}\cdot\left[\gamma_{jj}\cdot \ell n \underline{w}_j + \sum_{k=1}^{n}\gamma_{jk}\cdot \ell n \underline{w}_k + \gamma_{jt}\cdot \underline{t}\right].$$

Symmetry among the second order partial derivatives of C with respect to the input prices and linear homogeneity of C in $\boldsymbol{w}$ imply the following constraints

$$\gamma_{jk} = \gamma_{kj}, \quad \Sigma_{j=1}^{n}\alpha_j = 1, \quad \Sigma_{j=1}^{n}\gamma_{jk} = \Sigma_{k=1}^{n}\gamma_{jk} = \Sigma_{j=1}^{n}\gamma_{jt} = 0.$$

As in the case when output data are available these restrictions can all be tested.

Application of (12) – (12') to (21) and (22) gives the effect of technical change on the *TFP* growth rate according to

$$\nu \equiv \frac{\partial \ell n C}{\partial t} = -(\alpha_t + \gamma_{tt}\cdot \underline{t}) - \tfrac{1}{2}\cdot(\sum_{j=1}^{n}\gamma_{jt}\cdot \ell n \underline{w}_j), \tag{35}$$

where the first and second terms correspond to effects from neutral and non-neutral technical change, respectively. Further, by combining (11), (13) and (35)

$$T\hat{F}P = \nu \cdot \varepsilon^{-1} = -(\alpha_t + \gamma_{tt}\cdot \underline{t} + \tfrac{1}{2}\cdot \sum_{i=1}^{n}\gamma_{it}\cdot \ell n \underline{w}_i)\cdot \kappa \cdot \frac{C}{V}. \tag{36}$$

Since $T\hat{F}P$ is dependent on κ it will be necessary to perform a sensitivity analysis on $T\hat{F}P$ with respect to this parameter, unless it is known *a priori.*

Finally, it should be noted that in addition to the productivity measures the empirical analysis also yields estimates of (the logarithms of) the output quantity and output price indices. (Of course, like the estimates of *TFP* growth these estimates will be conditional on the markup factor κ.) By means of the definition (23a) and the results (24) and (34), the log of the output quantity index can be estimated according to

$$\ell n\,(y/y_o) = \underline{q}\cdot[\,(V/C)_o\cdot \kappa^{-3}\cdot \alpha_q^{-1}]\,. \tag{37}$$

Further, the definition

$$\ell n\,(V/V_o) \equiv \ell n\,(py/p_o y_o) \equiv \ell n\,(p/p_o) + \ell n\,(y/y_o)\,,$$

implies that the log of the output price index can be estimated as

$$\ln(p/p_o) = \ln(V/V_o) - \underline{q}\cdot[\,(V/C)_o\cdot\kappa^{-3}\cdot\alpha_q^{-1}]\,. \tag{38}$$

Since for many service industries proper output quantity and output price indices are not available in the national accounts statistics, (37) and (38) are important by-products of the estimation. For instance, in the Swedish national accounts quantity indices for the banking and the insurance industry are obtained by means of the *ad hoc* assumption that average labor productivity increases 2% per annum. By comparing the estimated indices (37) and (38) with the corresponding national accounts indices one can examine the empirical validity of this assumption.

V. Concluding comments

The problem considered in this paper concerns the possibilities to to empirically characterize a production process when there is complete input information but the output information is limited to data on the gross value of output, a situation typical of a large part of the private service sector. It is demonstrated that if (i) the technology is homothetic, if (ii) output can be treated as a scalar, and if (iii) the elasticity of total cost with respect output is strictly increasing in output, then, essentially, the only additional information required for a complete characterization of the production process is the possible difference (in percentage terms) between the marginal cost and the output price, i.e. the potential markup. Since in many cases it is difficult to obtain information about the markup the analysis proceeds to the case when the price elasticity is unknown but constant. It is shown that in this case the results continue to hold, *conditional* on the unknown markup factor.

The key assumption is (iii); it is this assumption that makes it possible to substitute known variables for the unknown output variable in the cost function.

The fact that (iii) is not only sufficient to enable this substitution but also necessary has an important implication, namely that a value measure of output carries information in excess to that inherent in input data only if the underlying technology is not homogeneous. Thus, that the technology exhibits non – constant returns to scale is a necessary but not sufficient condition.

Concerning productivity measurement, the result is that if the markup is known the rate of growth in total factor productivity can be estimated with the same precision as if output data were available. If the markup is unknown, the estimated rate of growth in *TFP* will be conditional upon the assumption made about the markup and so, in applications, it will be necessary to perform a sensitivity analysis where the effect of variations in the markup is assessed. This is, however, very easy to do; as long as different constant markups are considered the model does not have to reestimated when the markup is altered. Moreover, in quite a few empirical applications it should be possible to obtain information at least about the magnitude of the markup, indicating the interval over which the sensitivity analysis should be carried out.

Appendix : Conditions for uniqueness and stability of equilibrium

Since the profit functions are assumed to be twice differentiable with respect the y_j's, Theorem 2.6 in Friedman (1986, p. 45) can be used to formulate conditions under which the equilibrium is unique. According to this theorem, the equilibrium is unique if the symmetric $m \times m$ matrix

$$\mathbf{M} = \mathbf{P} + \mathbf{P}'$$

is negative definite, where $\mathbf{P}$ is the Jacobian matrix of the system of first order derivatives of the profit functions, i.e.

$$\mathbf{P} = \left[\frac{\partial^2 \pi_i}{\partial y_i \partial y_k}\right] .$$

The typical elements of $\mathbf{P}$ are:

$$\frac{\partial^2 \pi_i}{\partial y_i \partial y_k} = p''(1^t \boldsymbol{y}) \cdot y_i + 2 \cdot p'(1^t \boldsymbol{y}) \equiv A_i , \quad i \neq k \qquad \text{(A1)}$$

$$\frac{\partial^2 \pi_i}{\partial y_i^2} = A_i - f_i''(y_i) \cdot C_i(1,w,t) \equiv A_i - B_i \tag{A2}$$

implying that

$$\mathbf{M} = \begin{bmatrix} 2(A_1 - B_1) & A_1 + A_2 & \cdots & A_1 + A_m \\ A_2 + A_1 & 2(A_2 - B_2) & \cdots & A_2 + A_m \\ \vdots & & \ddots & \vdots \\ A_m + A_1 & A_m + A_2 & \cdots & 2(A_m - B_m) \end{bmatrix}$$

Denote the principal minor subdeterminants of $\mathbf{M}$ by $D_1, D_2, \ldots D_n$ (where, of course $D_m = |\mathbf{M}|$). For $\mathbf{M}$ to be negative definite, the D_j's should alternate in sign, starting with D_1 negative. As the number of firms (i.e. m) grows it becomes exceedingly more difficult to formulate simple conditions which guarantee that the subdeterminants obey these constraints. For this reason, only the cases where $m = 1$, $m = 2$ and $m = 3$ will be considered here.

By the concavity of the profit function the first subdeterminant is always (strictly) negative. The second subdeterminant can be written

$$D_2 = 4[(A_1 - B_1)(A_2 - B_2) - A_1A_2] - (A_1 - A_2)^2$$

which should be positive. A sufficient (but not necessary) condition for $D_2 > 0$ is that the inverse industry market demand curve is linear. Then $A_1 = A_2 = A < 0$ implying that the first term is strictly positive [since the B_i's are strictly positive by (4)], while the second term is zero.

Finally, to simplify the analysis when $m = 3$ only the case with a linear demand curve, i.e. $A_1 = A_2 = A_3 = A$, will be considered. In this case, one obtains, after a number of tedious manipulations, the following expression

$$D_3 = -2 \cdot A^3 + 6\bar{B}_a \cdot A^2 + 18(\bar{B}_g^3 / \bar{B}_h) \cdot A - 6\bar{B}_g^3$$

where $\bar{B}_a$, $\bar{B}_g$, and $\bar{B}_h$ denote the arithmetic, geometric, and harmonic means of B_1, B_2, and B_3, respectively. Further, since $\bar{B}_a \geq \bar{B}_g \geq \bar{B}_h$ (the inequalities being strict unless $B_1 = B_2 = B_3$)

$$D_3 \leq -2 \cdot A^3 + (6 + \epsilon)\bar{B}_g \cdot A^2 + 18\bar{B}_g^2 \cdot A - 6\bar{B}_g^3, \tag{A3}$$

where $\epsilon = (\bar{B}_a - \bar{B}_g) / \bar{B}_g$. It can easily be verified that if $\epsilon \leq 0.5$, which seems like a very reasonable assumption, then the RHS of (A3) will be non – positive for all A such that $-2 \cdot \bar{B}_g \leq A\ (< 0)$. (For higher values on ϵ the lower bound will be closer to zero.) Thus, for the case when $m = 3$ it should be possible to conclude that the equilibrium is unique if the inverse industry market demand curve is linear and (the absolute value of) its slope is less than twice the geometric mean of the slopes of the firm's marginal cost curves.

Stability conditions can be found in Friedman (1977, p. 71). According to these, the equilibrium is stable if

$$A_i - B_i + |\Sigma_{k \neq i} A_k| < 0 \qquad i = 1,\ldots,m. \tag{A4}$$

If the inverse demand curve is linear, then (A4) reduces to

$$(1-m)\cdot p'(1^t\boldsymbol{y}) - f_i''(y_i)\cdot C_i(1,\boldsymbol{w},t) < 0 \quad \forall\, i$$

i.e. for $m = 2$ the slope of the marginal cost curve should exceed the absolute value of the slope of the demand curve for each firm. If $m = 3$ the slopes of the marginal cost curves must be more than twice the absolute value of the slope of the demand curve. These conditions are considerably stronger than those required for uniqueness and fulfillment of them implies fulfillment of the uniqueness conditions.

REFERENCES

Appelbaum, E.: The estimation of the degree of oligopoly power. Journal of Econometrics 19, 287-299, 1982.

Binswanger, H.P.: The measurement of technical change biases with many factors of production. American Economic Review 64, 964-976, 1974.

Christensen, L.R., D.W. Jorgenson and L.J. Lau: Transcendental logarithmic production frontiers. Review of Economics and Statistics 55, 28-45, 1973.

Diewert, W.E.: An application of the Shephard duality theorem: A generalized Leontief production function. Journal of Political Economy 79, 481-507, 1971.

Friedman, J.W.: Oligopoly and the Theory of Games. North-Holland, Amsterdam, 1977.

Friedman, J.W.: Game Theory with Applications to Economics. Oxford University Press, New York, 1986.

Guilkey , D.K. and C.A.K. Lovell: On the flexibility of the translog approximation. International Economic Review 21, 137-147, 1980.

Mellander, E. and B.– C. Ysander: On the econometric analysis of production when there are no output data. Revised version of paper presented at the 1989 European meeting of the econometric society, 1990.

Morrison, C.J.: Markups in U.S. and Japanese manufacturing: A short – run econometric analysis. Journal of Business & Economic Statistics 10, 51-63, 1992.

Ohta, M.: A note on the duality between production and cost functions: Rate of returns to scale and rate of technical progress. Economic Studies Quarterly 25, 63-65, 1975.

Shephard, R.W.: Theory of Cost and Production Functions. Princeton University

Press, Princeton, New Jersey, 1970.

Zellner, A. and N.S. Revankar: Generalized production functions. Review of Economic Studies XXXVI, 241-250, 1969.