USING DETERMINISTIC AND STOCHASTIC DISTANCE FUNCTIONS TO MEASURE THE EFFECTS OF POLLUTION CONTROL ON A FIRM

Lauri Hetemäki

The Finnish Forest Research Institute, Research Papers 549
(Metsäntutkimuslaitoksen tiedonantoja 549)

Helsinki Research Centre
Helsinki 1994

USING DETERMINISTIC AND STOCHASTIC DISTANCE FUNCTIONS TO MEASURE THE EFFECTS OF POLLUTION CONTROL ON A FIRM

Lauri Hetemäki

The Finnish Forest Research Institute, Research Papers 549
(Metsäntutkimuslaitoksen tiedonantoja 549)

Helsinki Research Centre
Helsinki 1994

Hetemäki, L. 1995. Using deterministic and stochastic distance functions to measure the effects of pollution control on a firm. The Finnish Forest Research Institute, Research Papers 549. ISBN 951-40-1422-7, ISSN 0358-4283.*

The use of distance functions has recently become popular in production theory literature. Most of the empirical applications of distance functions have been based on the deterministic linear programming approach; few econometric studies exist. The purpose of this study is to provide new evidence on the performance of different methods in estimating output distance functions. In particular, results from deterministic and stochastic approaches and from using different functional forms of distance functions are compared. The methods are applied to study the impact of water pollution control on eight pulp plants in Finland. The results show that both approaches produce positive shadow prices for the water pollution variable. However, the results from the stochastic model are more stable. Also, the results stress the importance of carrying out sensitivity analysis when using the deterministic approach.

Keywords: *distance function, shadow prices of undesirable outputs, Finnish pulp industry, panel data*

Publisher: Finnish Forest Research Institute; project HE 3013
Approved by: Aarne Reunala
Authors's address: The Finnsih Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland (internet: hetemaki@metla.fi)

*I am grateful to helpful comments offered by Pekka Ilmakunnas and Jari Kuuluvainen. Skilled research assistance was provided by Mika Hakuni. This research was supported by funds provided by the Academy of Finland and Nordic Environmental Research Program (NERP).

1. Introduction

In recent years, there has been a growing interest in the use of distance functions in production theory.[1] The pioneering theoretical work on distance functions in production theory dates back to Shephard (1953, 1970) and the recent extensions can be found e.g. in Färe, Grosskopf and Lovell (1994), Färe and Grosskopf (1994), and Färe and Primont (1995). However, it is only in the last few years that empirical applications of distance functions have become more widespread.

The flexibility and generality of the distance function as an analytical tool is demonstrated in the number of different applications. For example, the studies showing how to use either input or output distance functions to measure technical or allocative efficiency or productivity are already numerous (e.g. Caves et al. 1982 a,b, Chavas and Cox 1994, Färe et al. 1989, Färe et al. 1990, Färe et al. 1994c). Chambers et al. (1994, p.1) showed that "the distance function is the unifying notion which links efficiency measures, quantity indexes and productivity indexes." Studies using distance functions to compute shadow prices of either inputs or outputs in regulated industries or services include e.g. Coggins and Swinton (1994), Grosskopf and Hayes (1993), Färe et al. (1993), Althin (1994), Hetemäki (1994 a,b). Furthermore, Lovell at al. (1990) demonstrate how to use distance functions to study income distribution and quality of life. Indeed, there are a number of factors (see below) which suggest that empirical applications of distance functions will increase rapidly in the near future.

In empirical applications, the great virtue of input and output distance functions is that they readily model multiple output production technologies and do not necessarily require price data to compute the parameters; only quantity data is needed. This is especially useful e.g. in production processes producing outputs, of which some are "bads" (pollutants) and do not have market prices. Further, distance functions do not impose any behavioural hypotheses (such as profit maximization or cost minimization) and allow production units to operate below the production frontier (i.e. to be inefficient). The latter property has proved to be particularly useful for the study of regulated industries and public services (such as hospitals and police). Finally, the duality results between the distance functions and the more conventional cost, profit and revenue functions provide flexibility for various empirical applications (Färe and Primont 1995).

Most of the empirical applications of distance functions have been based on deterministic nonparametric or parametric linear programming, and very few econometric studies exist. For example, the bulk of the applications in which the derivative properties of the distance function have been used for deriving shadow prices for inputs or outputs have been based on the translog linear programming model (e.g. Althin 1994, Clement, Grosskopf and Valdmanis 1994, Coggins and Swinton 1994, Färe et al. 1993). The deterministic linear programming approach does not require any distributional assumptions, is relatively easy to use and, in principle, allows for the computation of a large number of parameters even with a small number of observations. The major weakness of the approach is that it does not allow random disturbances and provides no statistical criteria for the consistency of the results. Thus, in order to justify the approach, one has to assume that measurement errors can be neglected or that they are all of the same sign (negative). Moreover, the efficiency of these estimators is an open question, since expressions for their asymptotic covariance matrices have never been devised (Green 1993 b). On the other hand, the econometric approach allows for random disturbance and provides information about the statistical significance of the results, but at the cost of assuming a specific distribution for the error term. However, the relative merits of the deterministic and stochastic approaches is not only a theoretical issue but also an empirical one.

The comparison of the different approaches is hindered by the fact that there have been no empirical applications comparing the different approaches for computing distance functions. One purpose of this study is to provide new empirical evidence of the relative performance of deterministic and stochastic parametric distance functions.[2] In particular, the relative strengths of the parametric deterministic distance function (DDF) and the stochastic distance function (SDF) approaches are examined using a particular case study. A real world data set is used, rather than controlled or Monte Carlo data. This is because the purpose of the study is also to provide new information about the particular case involved.[3] The DDF and SDF approaches are applied to examine the impact of water pollution control on the production technology of the Finnish pulp industry and to derive a measure for the cost of reducing different water pollution effluents. The theoretical framework of the present study is based on Färe et al. (1993). In the empirical part of their study, Färe et al. analyse the effects of pollution control in the US pulp and paper sector using a deterministic parametric linear programming approach and plant level *cross section* data (30 plants). In the present study, the data is based on observations from eight pulp plants in Finland over a period of 19 years (1972-90), i.e. plant level balanced *panel data* is used. The models are estimated using both pooled data and panel data specifications. Furthermore, in order to see how sensitive the results are to the choice of functional specification, a restricted translog and a Cobb-Douglas function are used. Although, the primary purpose of the study is to analyse the comparative performance of the two methodologies and the different model specifications, the study also produces new results on the impacts of pollution control on plant revenue and performance. These results are interesting in that they can be interpreted to provide support for the recently widely discussed and controversial *"Porter hypothesis"* (Porter 1990, Oates et al. 1993).

The paper is organized as follows. Section 2 briefly sets out the concepts and theoretical framework. In section 3 the two different empirical approaches are outlined. Section 4 discusses the data and choice of empirical variables. Section 5 presents the empirical results and their implications. Concluding remarks are given in section 6.

2. Theoretical model

A production technology transforming factors of production $x = (x_1, x_2, ..., x_n) \in R_+^n$ into outputs $y = (y_1, y_2 ...., y_m) \in R_+^m$ can be modelled by the output set P(x). This set contains all technically feasible output vectors for the input vector x, i.e., P(x) = {y $\in R_+^m$: x can produce y}. It is assumed that the technology satisfies the axioms of an output distance function (e.g. Färe and Primont 1995, Chp. 2). In particular, outputs are assumed to be only *weakly disposable*. Conventionally, the assumption of *strong* (or free) *disposability* is made.[4] The output distance function is defined on the output set P(x) as

$$(2.1)\quad D_o(x, y) = \min_{\theta}\{\theta: (y/\theta) \in P(x)\}$$

Equation (2.1) gives the largest radial expansion of the output vector, for a given input vector, which is consistent with that output vector belonging to P(x). The axioms regarding the output set P(x) impose a set of properties on the output distance function (e.g. Färe and Primont 1995). The value of the output distance function must be less than or equal to one ($D_o \leq 1$) for any feasible output. Further, the value of the distance function is the reciprocal of the *Farrell output-based technical efficiency index* (Färe, Grosskopf, and Lovell 1994).

The revenue function defined by (Shephard 1970, Färe and Primont 1995),

$$(2.2)\quad R(x, r) = \max_{y}[ry: y \in P(x)]$$

can also completely describe the production technology, where the output price vector is denoted by r = $(r_1, ..., r_m)$ and it is assumed that r can be nonpositive. The revenue function describes the maximum revenue that can be obtained from the given technology at output prices r. Shephard (1970) showed that the revenue function and output distance function are dual. Consequently, we can define the *revenue function* in terms of the distance function and vice versa. Formally, (2.3a) $$R(x,r) = \max_{y}\{ry: D_o(x,y) \leq 1\}$$

(2.3b) $$D_o(x,y) = \max_{r}\{ry: R(x,r) \leq 1\}$$

Thus, the revenue function can be derived from the output distance function by "maximizing" revenue over output quantities and that the output distance function is obtained from the revenue function by maximizing over output prices.

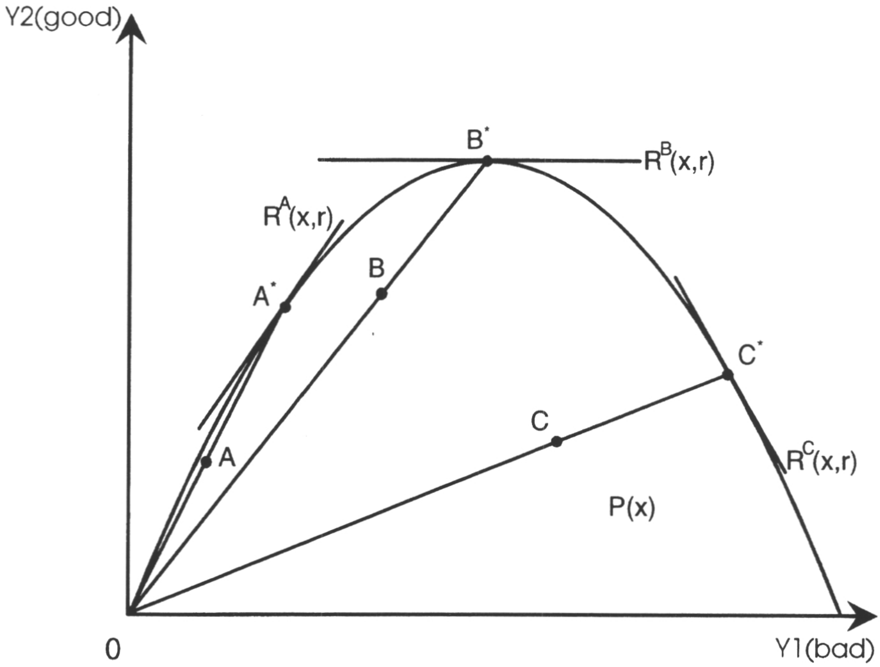
Following the analysis of Färe et al. (1993), it can be shown that the *revenue deflated* output shadow prices for each observation can be derived as the derivative of the distance function (using dual Shephard's lemma). These are *relative* output shadow prices. In order to obtain *absolute* (undeflated) shadow prices, additional information regarding the revenue is required. Färe et al. (1993) show that the *absolute* shadow prices can be computed when maximal revenue R(x, r) is known. The assumption which allows for the computation of the absolute shadow prices is (following Färe et. al. 1993): *One observed output price equals its absolute shadow price.* This assumption implies that at least one shadow output price equals its market price and it allows different plants to face different competitive markets. Alternatively, one could assume that in one output market observed revenue equals maximum revenue. Let output 1 denote the good output and assume that the observed good output price (r_1^o) equals its absolute shadow price (r_1^s), i.e. for m =1, $r_1^s = r_1^o$. The absolute shadow prices for each observation of undesirable outputs (m = 2,...,M) can now be computed as

(2.4) $$r_m^s = R^*(x,r^s) \bullet (\partial D_o(x,y)/\partial y_m) = r_1^o \bullet \frac{\partial D_o(x,y)/\partial y_m}{\partial D_o(x,y)/\partial y_1}$$

In equation (2.4) the ratio of output shadow prices reflects the relative opportunity cost of the outputs in terms of foregone revenue, i.e. it is equivalent to the marginal rate of transformation. It may be noted that the above expression does not require information on regulatory constraints. This is important because often data on regulations are not available, and even if such data is available, the plants rarely operate exactly at the level of the constraint. Moreover, there may be incentives for the firms to reduce pollution which are not related to the environmental regulation (see Porter 1990). Thus, "shadow prices reflect the tradeoff between desirable and undesirable outputs *at the actual mix of outputs,* which may or may not be consistent with the maximum allowable under regulation" (Färe et.al. 1993, p. 376).

The result of equation (2.4) is also illustrated in *Figure 1.*, in which the output set P(x) consistent with weak disposability of bad output (pollution) is shown. Consider three possible points: A, B and C. Each of them are below the frontier (i.e. production is technically inefficient), and one cannot compute the shadow prices at these points, since there is no tangent hyper plane to support the points. Instead, the shadow prices have to be calculated "as if" they were on the boundary. The inefficient points are radially (proportionally) scaled up to hypothetical observations on the frontier (points A*, B*, C*). By definition, the output distance function seeks such a scaling. Therefore, the derivatives of equation (2.4) can be calculated for the observed inefficient points and they yield the same mutual relation as the derivatives for the hypothetical observation, since radial scaling does not affect the shadow price relation. At point A*, the shadow price of bad is negative, at point B* it is zero, and at point C* positive. The plant may be operating at point A*, because of environmental regulation or "constraints" originating from consumers preferences ("green values"). At point B*, the bads do not affect plant revenue. Finally, point C* may be possible if, for example, the measures which increase the productivity of the production process also reduce bads, e.g. by reducing material waste and/or saving energy (c.f. *Porter hypothesis*).

Figure 1. Technology P(x) and revenue R(x,r) with weak disposability of Y1



3. Empirical Models

The output distance function can be "estimated" in several ways; see e.g. Hetemäki (1994 a,b), Grosskopf and Hayes (1993) and Lovell et al. (1990). First, there is the choice between a nonparametric and parametric model. For the present paper the nonparametric model is excluded because it is piecewise linear and thus not differentiable (especially at the corners) and therefore the shadow price parameters cannot be determined directly.[5] The second choice is between deterministic and stochastic parametric models. Below both of these types of models are presented. Further, one has to choose from a number of different functional forms and estimation methods.

3.1 Parametric linear programming model

In order to estimate the deterministic output distance function, a parametric functional form is defined. Following the earlier studies in this literature (e.g. Althin 1994, Färe et al. 1993), the translog function is used. As is well known, the advantage of this form is its flexibility. Moreover, it does not impose strong disposability of outputs.

$$
\begin{aligned}
(3.1)\qquad \ln D_o(x,v,w) = {} & \alpha_0 + \sum_{i=1}^{m} \alpha_i \ln v_i + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i + 1/2 \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} (\ln v_i)(\ln v_j) \\
& + 1/2 \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} (\ln w_i)(\ln w_j) + 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} (\ln x_i)(\ln x_j) \\
& + \sum_{i=1}^{m} \sum_{j=1}^{M} \mu_{ij} (\ln v_i)(\ln w_j) + \sum_{i=1}^{m} \sum_{j=1}^{n} \eta_{ij} (\ln v_i)(\ln x_j) \\
& + \sum_{i=1}^{M} \sum_{j=1}^{n} \xi_{ij} (\ln w_i)(\ln x_j)
\end{aligned}
$$

In (3.1), $v = (v_1, \dots, v_m)$ denotes desirable (good) outputs, $w = (w_1, \dots, w_M)$ undesirable (bad) outputs, and $x = (x_1, \dots, x_n)$ inputs. The following symmetry (S) and homogeneity (H1-H4) restrictions are imposed

$$
\begin{aligned}
(S)\qquad & \alpha_{ij} = \alpha_{ji}, \qquad \forall i,j = 1,\dots,m \\
& \beta_{ij} = \beta_{ji}, \qquad \forall i,j = 1,\dots,M \\
& \gamma_{ij} = \gamma_{ji}, \qquad \forall i,j = 1,\dots,n
\end{aligned}
$$

(H1) $\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{M} \beta_j = 1$

(H2) $\sum_{j=1}^{m} \alpha_{ij} + \sum_{j=1}^{M} \mu_{ij} = 0 \qquad \forall i = 1,\ldots,m$

(H3) $\sum_{j=1}^{M} \beta_{ij} + \sum_{j=1}^{m} \mu_{ji} = 0 \qquad \forall i = 1,\ldots,M$

(H4) $\sum_{j=1}^{m} \eta_{ji} + \sum_{j=1}^{M} \xi_{ji} = 0 \qquad \forall i = 1,\ldots,n$

The parameters of equation (3.1) are computed using the linear (or goal) programming formulation suggested by Aigner and Chu (1968). From the theory (Chp. 2) it is known that for each observation the value of the distance function must be less than or equal to 1, i.e. $\ln D_o$ must be less than or equal to zero (assuming there are no measurement errors). Formally,

(3.2) $\ln D_o^K(x, v, w) \leq 0 \qquad \forall k = 1,\ldots,K$

where K denotes the observation. By adding a non-negative "error" term, one can rewrite (3.2) as

(3.3) $\ln D_o^K(x, v, w) + \varepsilon^k = 0$

where $\varepsilon(\varepsilon \geq 0)$ denotes the error term. It may be noted that it is customary in the literature to interpret the non-negative "error" term as the reciprocal of the Farrell output-based technical efficiency index. Next we choose the "fitting" criteria to be the *minimum absolute error (MAE)* criteria, i.e. $\sum_{k=1}^{K} \left|\varepsilon^k\right|, \quad \varepsilon^k \geq 0$. MAE fits $\ln D_o$ so that the sum of errors is as small as possible.

The complete parametric linear programming problem, with restrictions, can be expressed as

$$(3.4)\quad \max \sum_{k=1}^{K} \left[\ln D_o(x^k, v^k, w^k - \ln 1 \right]$$

s.t.

$$\text{(i)}\quad \ln D_o^k(x^k, v^k, w^k) \leq 0 \qquad k = 1, \ldots, K$$

$$\text{(ii)}\quad \frac{\partial \ln D_o(x^k, v^k, w^{k)}}{\partial \ln x_i} \leq 0$$

$$\text{(iii)}\quad \frac{\partial \ln D_o(x^k, v^k, w^{k)}}{\partial \ln v_i} \geq 0$$

In addition to the above restrictions, the symmetry restrictions (S) and homogeneity restrictions (H1-H4) are imposed. The homogeneity constraint also ensures that the technology satisfies weak disposability of outputs. The constraint (i) restricts observations to be on or below the frontier technology. The constraint (ii) imposes strong disposability of inputs, i.e. increasing inputs can never reduce outputs (congestion is not allowed). Finally, constraint (iii) ensures that the desirable output shadow prices are greater than or equal to 0.

It may be noted that compared to the model of Färe et al. (1993), model (3.4) differs in three respects. First, in the present study no restrictions are imposed on the shadow prices of undesirable outputs, whereas Färe et al. set the shadow prices of undesirable outputs to be negative or zero ($\partial \ln D_o(x^k, v^k, w^k) / \partial \ln w_i^k \leq 0$). Secondly, Färe et al. do not impose constraint (ii) above, i.e. they allow input congestion. Thirdly, equation (3.4) is estimated as a panel data model (Färe et al. 1993 use cross section data with 30 observations). The panel data model in this context is called by Lovell (1993) the "inter temporal goal programming approach". This approach has been used by Nishimizu and Page (1982) in a single-output production frontier context. In principle, the approach allows the calculation of technical efficiency and shadow prices for each producer in each time period, as well as the shift in the production frontier over time (technical change).

3.2 Econometric model

In order to transform the deterministic equation (3.1) into a stochastic one, a random disturbance term has to be added. Random disturbances may arise, e.g. because of measurement errors in the data, randomness of the efficiency distribution between the plants over time, luck etc. In the present study, an error term (ε^k), assumed to be independently and identically distributed as $N(0,\sigma_\varepsilon^2)$, was added to equation (3.1). The estimation of a stochastic distance function is not as straightforward as, e.g. the estimation of a production, cost or profit function. The basic problem with distance functions, as concerns econometric estimation, is that one does not usually observe (have data on) the dependent variable. Further, if one sets the distance function equal to its efficient (frontier) value, $D_O = 1$, the left-hand side of the distance function is invariant, an intercept cannot be estimated, and OLS parameter estimates will be biased. Further, if the distance function is expressed in logarithms, the left-hand side of the distance function will be zero for all observations (i.e. D = ln(1) = 0).

In the present study, the approach of Grosskopf et al. (1992) and Lovell et al. (1990) is used to estimate the stochastic distance function. This procedure imposes the value of 1 on the distance function and uses the homogeneity property (i.e. outputs homogeneous of degree +1) to solve the invariance problem of the left-hand side of the distance function. Thus, the procedure amounts to estimating the technology frontier, i.e. it is assumed that the fitted values of the distance function deviate from 1 only due to stochastic error term. Homogeneity is imposed by multiplying all output values on the right-hand side and the value of the distance function on the left-hand side by a numeraire variable. This transformation causes the multiplicative variable to appear on both sides of the equations, which may result in endogeneity on the right-hand side. Therefore one has to test whether the errors are correlated with the regressors, and if this is the case, use the instrumental variables estimation method.

In order to transform equation (3.1) to estimable form, the dependent variable and the output terms of equation (3.1) were multiplied by $\lambda = 1/SS$, where SS is the amount of suspended solids in the waste water (i.e., the transformation is first computed in levels form, after which the logarithmic transformation is taken). This transformation imposes the homogeneity of outputs restriction and weak disposability of outputs.

Finally, it may be noted that although the distance function can be used to study effects such as elasticities of substitution, economies of scale, and technical change, the primary concern here is the shadow prices of undesirable outputs.

4. Data and variables

The empirical analysis is based on data from the Finnish sulphate pulp industry. The institutional background and the water pollution regulations concerning this sector have been discussed in more detail in Hetemäki (1994 a,b.). It suffices here to say, that sulphate pulp mills are usually classified as integrated pulp and paper plants or non-integrated pulp plants. The first group consists of plants in which the production process is integrated with the production of paper or paperboard and the latter group comprises plants that produce only sulphate pulp (to export or to sell to domestic paper plants). In 1990 there were altogether 17 sulphate pulp plants in Finland, of which 7 were non-integrated. The sulphate pulp industry represents a typical process industry, whose inputs and end products are relatively homogeneous in comparison with most other industries. Thus, the inputs and outputs are also relatively accurately measurable. A major part of the output is used domestically; of total output, exports were 34 % in 1972, 38 % in 1980 and 26 % in 1990. However, of the end product (paper/paperboard) approximately 90 % is exported. The main water effluents produced jointly with pulp are biological oxygen demand (BOD), suspended solids (SS), nitrogen (N), phosphorous (P), chemical oxygen demand (COD), and absorbable organic halogens (AOX).

In order to keep the sample as homogeneous as possible, the empirical analysis includes only those plants which were operating during the whole period studied. By this procedure the bias involved in comparing plants with different vintages of production technology is reduced, although not totally removed. The data set used in the empirical analysis is based on a balanced panel containing annual data from 8 sulphate pulp plants over the period 1972-90. All the plants are non-integrated, except one, for which it was possible to separate the sulphate pulp production from the paper/paperboard production (in terms of the data needed). The plants in the sample have accounted for more than a half of the total production of the sulphate pulp industry during 1972-90.

The data used for the estimations consist of observations on quantity (Q) and gross value (GVP) of sulphate pulp output, net fixed capital stock (K), hours worked (L), value of materials input (M), biological oxygen demand (BOD), total waste water flow (FL), and suspended solids (SS) (see Appendix I). Although, the waste water flow has not been regulated by the water authority, its reduction has nevertheless been perhaps the most important means by which the plants have tried to reduce different water pollution substances. Moreover, the constraint need not originate from the regulating authority but may come from the consumer preferences. This has recently been evidenced by the "green marketing" strategies of pulp and paper manufacturers.

The standard deviations, means, minimum and maximum values, skewness and kurtuosis are shown in *Table1.* The standard deviations for all the variables are less than their mean values, indicating that the mills are a relatively homogeneous group.

Table 1. Descriptive Statistics. 8 sulphate pulp mills observed annually between 1972 - 1990 (Sample Size = 152)

VARIABLE	UNIT	MEAN	ST. DEV	MIN	MAX
Q	1000 t	224.9	94.3	87.3	511.8
GVP	mill. FIM	602.9	231	209.7	1200
M	mill.FIM	420.2	165.4	137.6	848
L	1000 h	811.4	353.1	209.7	1803
K	mill. FIM	974.2	372.3	273.2	1797
FL	mill. m3	42.1	20.4	15.2	126.7
SS	t	1978.3	1569.3	277	9950
BOD	1000 t	6034.8	2988	554	15370

Q = pulp output; GVP = gross value of output in 1990 prices; M = value of materials input in 1990 prices; L = hours worked (productive and non-productive workers); K = net fixed capital stock in 1990 prices; FL = waste water flow; SS = suspended solids; BOD = biological oxygen demand.

5. Empirical results

The initial estimation results showed that the complete translog model could not be estimated for the stochastic model due to multicollinearity (singularity of the Hessian matrix). Although, the deterministic approach allowed the computation of this model specification, the "fitted" values were all equal to 1 and the parameter values were extremely sensitive even to minor changes in model specification or data.[6] Thus, it was thought to be more appropriate to compare functional forms, which could be regarded as not suffering from these biases. Consequently, the empirical results are based on two functional forms: (i) a special case of a translog function, with a first-order approximation in the input quantities and second -order terms in the output quantities (eq. 5.1), and (ii) a Cobb-Douglas form (eq. 5.2).

$$(5.1) \qquad \ln D_o(x,v,w) = \alpha_0 + \alpha_p + \alpha_t + \sum_{i=1}^{m} \alpha_i \ln v_i + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i + 1/2 \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} (\ln v_i)(\ln v_j)$$
$$+ 1/2 \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} (\ln w_i w_j) + \sum_{i=1}^{m} \sum_{j=1}^{M} \mu_{ij} (\ln v_i)(\ln w_j)$$

$$(5.2) \qquad \ln D_o(x,v,w) = \alpha_0 + \alpha_p + \alpha_t + \sum_{i=1}^{m} \alpha_i \ln v_i + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i$$

where x denotes inputs (capital (K), labor (L), and materials (M)); v denotes desirable outputs (quantity of pulp produced (Q)), and w undesirable outputs (biological oxygen demand (BOD) and waste water flow (FL)). In some model specifications, the plant-specific (α_p) and period-specific (α_t) fixed or random effects were included (see below). The plant and time subscripts are deleted for simplicity from eqs. (5.1-5.2); the observations run across plants (p = 1,....,8) and over time (t = 1,...., 19), i.e. for a total of 152 observations. Finally, in the stochastic model, an error term was added to equations (5.1) and (5.2) and the dependent variable and the output terms were transformed in order to impose homogeneity and allow for the estimation (see above).

Since, a priori, there is usually not enough information which enables one to choose a stochastic specification which approximates the *data generating mechanism* the best, a number of different specifications has to be estimated. In present study, the stochastic model was estimated using the following specifications: pooled data without plant- and period-specific effects, and five different panel data specifications (one- and two-way fixed and random effects models and a random coefficients model). The deterministic linear programming model was computed both for the pooled and fixed effects specifications. To test whether plant- and period-specific effects are present, we employed an F-test for the fixed effects model and the Lagrange multiplier test of Breusch and Pagan for the random effects model. Moreover, the Hausman test was run to test whether the fixed or random effects specification should be used. Finally, a chi-squared test of the random coefficients model against the alternative of the classical regression (no randomness in the coefficients) was computed.[7] Of the different model specifications, the two factor fixed effects (or covariance) model (TFFE) proved to be preferable in terms of model diagnostics (see *Table 5.*). However, the results showed that the fixed and random effects models (but not the random coefficients model) produce rather similar estimates. Consequently, whether the individual- and time-specific effects are treated as fixed or random does not change the results significantly.

In the TFFE specification, the intercept is allowed to vary from plant to plant and period to period, while the slope parameters are assumed to be constant over both plants and time periods. The plant- and time-specific effects are typically assumed to arise from the omission of variables whose explicit inclusion in the model is not possible. For example, factors such as environmental regulations, management and infrastructure differ across plants and may affect the efficiency of the plants. These affects are captured by adding a dummy variable for each plant (PL2-PL8). Also, it is likely that plants are affected by, e.g. macroeconomic factors (oil price shocks) and general environmental attitudes of society, which vary over time. The latter effects are captured by the time dummies that vary over time but not over plants (Y73-Y90), i.e. it is assumed that similar factors "hit" every plant in each time period. In the TFFE model the plant- and time-specific factors are allowed to be correlated with inputs and outputs, and the model is estimated consistently by OLS.

The results reported in *Table 2.* show that in the stochastic models the measures of goodness of fit ($\bar{R}^2$) were high. However, the parameter restriction test rejected the Cobb-Douglas specification. Also, the Cobb-Douglas and translog model without fixed effects suffered from various specification problems (see Appendix II).

The shadow prices were computed only for waste water flow in the stochastic model, since the parameters used for computing the shadow price for BOD were either close to zero or insignificant. However, it should be added, that the shadow prices of FL in the stochastic models were not sensitive to the exclusion of BOD from the models (for the deterministic model the shadow prices for BOD were computed for illustrative purposes, see *Table 3*). The most significant feature of the shadow price estimates, shown in *Table 3.* is that for bulk of the observations, the FL shadow prices are positive. It should be stressed that this does not mean that the effect of *environmental regulation* is positive. The output distance function measures the effect of *pollution control*, not the effect of regulation. The regulation is always an exogenous restriction on the production, but control of pollution may be a free choice of the firm for various reasons (see Porter).

In the restricted translog two-way fixed effects models, 3 out of the 152 observations were negative when stochastic approach was used, and 9 observations were negative when deterministic approach was used. In terms of *Figure 1.*, these results would imply that the plants have most of the time been operating around the point C* (either on the frontier or below it). The positive shadow prices for FLOW are probably due to the fact that the internal process changes in the production of pulp have simultaneously decreased the amount of waste water effluents and improved productivity. The long-run strategy of the pulp plants in developing the production process has been to aim at closed-loop water systems, which simultaneously improve efficiency in the control of production systems and reduce water pollution. As a result of this strategy, the production of one ton of pulp in 1990 generated on average three times less waste water than in the 1972.

It is important to note that the pollution control may have occurred either independently of the regulations or as a result of the regulation (or due to both of these). In the first case, waste water reduction may have emerged as a by-product of productivity improvement measures or "green marketing strategies". In the latter case, regulation forces firms to adjust production process and reduce pollution, which, however, may also lead to improvements in productivity. Thus, there may be potentially significant *"learning by doing"* effects associated with environmental regulations. This type of argument has recently been put forward by

Table 2. Parameter estimates (152 obs.)

	Cobb-Douglas		Cobb-Douglas FE		Restricted Translog FE	
	deter.	stoch.	deter.	stoch.	deter.	stoch.
constant	-0.94	0.03 **	-0.51	0.01	-0.21	-0.02
K	-0.28	-0.08	-0.19	-0.14 **	-0.20	-0.15 **
L	0.00	0.06 *	-0.20	-0.16 **	-0.23	-0.17 **
M	-0.58	-0.75 **	-0.50	-0.67 **	-0.48	-0.66 **
Q	0.78	0.88 **	0.90	0.88 **	0.83	0.89 **
BOD	-0.03	0.05 **	-0.03	-0.02	-0.18	-0.01
FL	0.24	0.06	0.13	0.13 **	0.36	0.09 *
Q2					0.12	0.06 *
BOD2					0.02	-0.01
FL2					0.31	0.04
BODFL					-0.10	0.02
BODQ					0.09	-0.01
FLQ					-0.20	-0.06 **
PL2			-0.21	-0.08 **	-0.22	-0.04
PL3			0.13	0.09 **	0.16	0.15 **
PL4			-0.07	-0.14 **	-0.07	-0.09
PL5			0.02	-0.04	0.05	-0.00
PL6			-0.03	-0.10 **	-0.03	-0.05
PL7			-0.24	-0.21 **	-0.29	-0.15 **
PL8			-0.41	-0.35 **	-0.41	-0.32 **
Y73			-0.05	-0.06 **	-0.05	-0.06 *
Y74			0.05	0.07 **	0.05	0.08 **
Y75			-0.05	0.26 **	-0.06	0.27 **
Y76			0.15	0.31 **	0.16	0.31 **
Y77			0.18	0.30 **	0.21	0.31 **
Y78			0.11	0.09 **	0.13	0.11 **
Y79			-0.16	-0.01	-0.09	0.01
Y80			-0.09	-0.01	-0.05	0.01
Y81			-0.08	0.01	0.00	0.03
Y82			0.10	0.15 **	0.17	0.16 **
Y83			-0.01	0.02	0.03	0.03
Y84			-0.08	-0.00	-0.05	0.00
Y85			-0.04	0.01	-0.01	0.02
Y86			-0.08	0.03	-0.06	0.02
Y87			-0.24	-0.03	-0.18	-0.04
Y88			-0.22	-0.08	-0.18	-0.09
Y89			-0.20	-0.07	-0.17	-0.08
Y90			-0.19	-0.06	-0.15	-0.05
adj. R2		0.95		0.98		0.98

** and * indicate the heteroskedasticity-consistent t-values of the parameters that are significant at the 5% and 10% level, respectively.

Table 3. Waste water flow shadow prices

Model	Mean shadow price, FL	Standard deviation	Minimum	Maximum
CD deter.	840.1	114.8	522.2	1176
CD stoch.	171.4	23.2	106.6	240
CDFE deter.	392.7	53.6	244	549.8
CDFE stoch.	406.5	55.5	252.7	569.2
RTRFE deter.	759.4	453.8	-269.3	2169
RTRFE stoch.	278.1	108.8	-33.14	585.9

CD = Cobb-Douglas, CDFE = Cobb-Douglas function with plant and period specific fixed effects, RTRFE = restricted translog function with plant and period specific fixed effects. It may be noted that the mean value for the computed BOD shadow price in the RTRFE deterministic model is FIM 1521.6 (min. -33.1 and max. 585.9).

Porter (1990). Indeed, in their analysis of the *"Porter hypothesis",* Oates et al. (1993) argue that the most likely reason that regulations might generate positive effects on firms' profits is that there has been inefficiency and unrealized opportunities for cost-savings and product enhancement before the regulation and that the regulations induce the realization of these opportunities.

If the above argument is correct, one may ask why the firms utilized the positive spillover effects only after the regulations forced them to do so. One possible answer is related to the information cost argument. There are many potentially ways by which production efficiency could be enhanced, and firms are uncertain about which of the possibilities will result in benefits that exceed the (research and development) costs. However, analyzing a wide range of possible new ways to increase productivity is costly, and the firms may not utilize these possibilities until regulations force them to do so. Oates et al. 1993 list a number of other possible reasons why firms do not realize the potential gains in the absence of regulation.

The results obtained in the present study are in accordance with those obtained in Hetemäki (1994b), using the same theoretical framework and data set but applying a "two-stage method" to estimate a stochastic distance function. In contrast, the present results are rather different from those obtained by Färe et al. (1993) using a deterministic output distance function and cross-section data for pulp mills operating in Michigan and Wisconsin in 1976. The results obtained by Färe et al. showed that the absolute shadow prices of different measures of water pollution were large and negative and that there are wide variations in shadow prices across the different mills. For example, the mean of the plant -specific absolute shadow price of BOD indicated that reducing one ton of BOD emissions diverts enough resources to have produced over two tons of paper, and the standard deviation of the plant -specific shadow prices was higher than their mean. If these differences in the results of the two studies could be regarded as reflecting purely country -specific differences, it would indicate that Finnish and US pulp mills are using very different production technologies or/and that they are operating in a strikingly different environment. However, it would seem more plausible to consider that a significant part of the difference is a result of differences in the data bases (homogeneous panel data vs. heterogeneous cross-section) and in the fact that the parameter restrictions differ in the two studies (e.g. Färe et al. impose negative shadow price). Indeed, the study by Evans and Heckman (1988) showed, in a different context, that the parametric linear programming approach is very sensitive to the constraints imposed on the technology.

5.1 Comparing the results

How do the estimates using the two different methods compare? If we simply compare the columns for deterministic specification with columns for the stochastic specifications in *Table 2.*, we can see that the differences in the coefficients between the three deterministic model specifications are larger than the respective differences between the stochastic models. That is, the computed coefficients of the linear programming models are more sensitive to small changes in specifications; in the stochastic models they are rather stable. For the stochastic approach, alternative model specifications lead to essentially the same conclusions concerning shadow prices, but for the deterministic approach minor changes in the specification (or data) lead to major changes in the inferences (see *Table 3*). One would usually prefer the most flexible a functional form, which in the present study is the translog. However, using the results from the stochastic model estimations, the multicollinearity problems associated with this specification became apparent; the data simply do not contain enough information to enable the identification of all the parameters in the translog model.

Since the model diagnostics for the stochastic specifications indicated that the restricted translog model with the fixed effects performed the best, it is interesting to examine more closely the differences between the two approaches for this specification. The shadow prices for waste water flow (FL) for the two approaches are shown in *Figures 2* and *3.* In *Figure 2* the mean shadow prices, measured on the y-axis, are given in Finnish marks (FIM) across the plants over time (1972-90). For both approaches, the shadow price follows a similar downward trend, except for the last two years (89-90), when the shadow price for the deterministic model increases. Also, the level of the shadow prices differs: it is on average 2.5 times higher for the deterministic model. In *Figure 3* the waste water flow shadow prices scaled by the respective mean values are shown for the whole sample; observations 1-19 show the values for plant 1, observations 20-38 show the values for plant 2, etc. The figure shows clearly that for the deterministic model, both the scale and frequency of variation in the shadow price is larger than for the stochastic model.

In summary, both approaches show the same general downward pattern over time for shadow prices and both are positive. However, in the deterministic model the level of the shadow price is much higher and the variation wider. The decreasing shadow prices probably reflect the general change in the production process towards a closed-loop water

Figure 2. Waste water flow shadow prices (FIM) from 1972-90

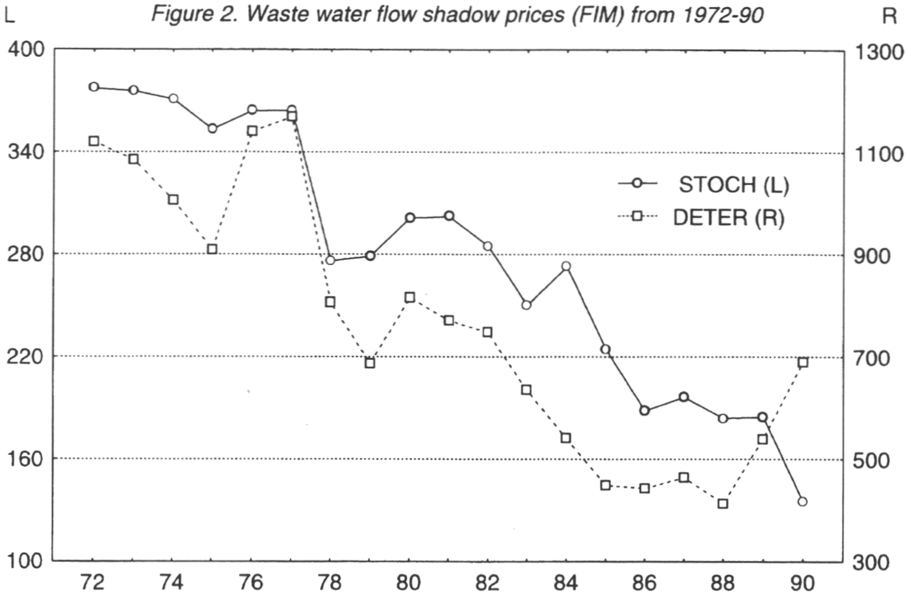
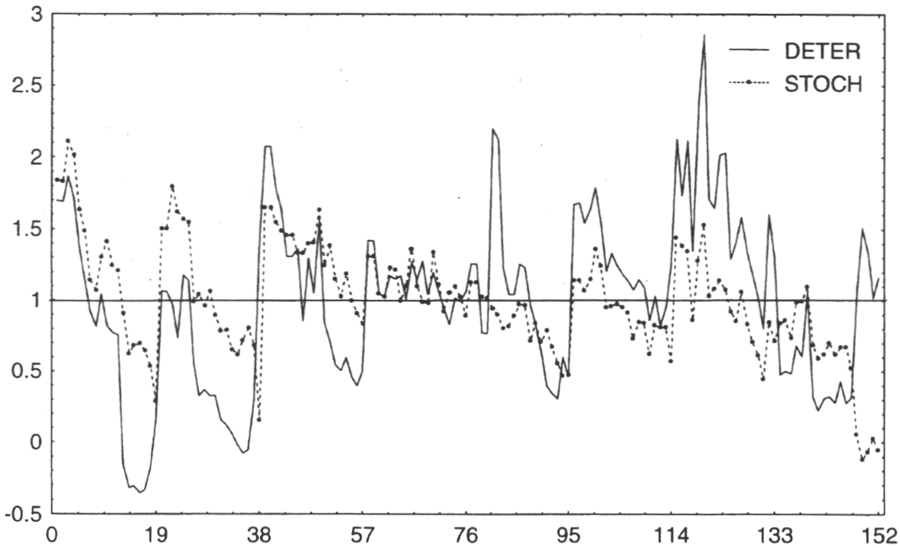


Figure 3. Waste water flow shadow prices scaled by mean values (obs. 1-152)



On the horizontal axis, the observations for each plant are shown, thus obs. 1-19 for plant 1, obs. 20-38 for plant 2,....., obs. 133-152 for plant 8.

system and the fact that the positive spillover effects (efficiency gains) associated with this process change get smaller at the margin (in terms of *Figure 1*, plants are moving from points like C* towards B* or A*). Against this evidence, it would seem unrealistic that the shadow prices for flow have increased significantly in 1989-90, as indicated by the deterministic model. Also, the absolute values of the shadow prices from the deterministic model appear unrealistically high (the maximum shadow price is FIM 2169, which is almost equal to the mean of the gross value of pulp output per ton FIM 2728).

Finally, there is an important question which arises with the use of the deterministic model. Namely, if only the deterministic approach had been used (as in the bulk of the previous studies), which model specification would we have ended up with? Probably the complete translog (because of its flexibility) as in so many of the previous studies. However, as the results indicated, for the data used in this study, the complete translog form suffered from serious multicollinearity, due to which the parameters could not be precisely identified (and the stochastic model not even estimated). Moreover, how could have we chosen whether to include or exclude the plant and time specific fixed effects in the model? These questions bring forward the important drawback with the parametric linear programming approach, i.e. the lack of tools to guide the model specification search.

6. Conclusions

The implications of the present study are of two kinds: those related to the effects of pollution control on the sulphate pulp plants and those related to methodological issues. The substantive implications are summarised first.

The results of the two approaches used are coherent in that both indicate that water pollution reduction by Finnish pulp plants has been, for most of the plants and for most of the period studied, associated with the increase in revenues. For some plants and some years, the effects have been either slightly negative or close to zero. However, the positive shadow prices should not be interpreted to show that environmental regulations cause plants' profits to increase. Rather, the result indicates that control of emissions is part of the control of the whole pulping process. Recycling waste water and closing the water circulation simultaneously reduces material waste, improves the production process and reduces water pollution. In other words, environmental regulation is not the only factor which has caused these plants to reduce water pollution, but also the fact that pollution control measures and improvements in the production process appear to be strongly positively correlated. How well the above result can be generalized to other production processes is an empirical issue. Nevertheless, the result indicates that one should not a priori rule out the possibility that pollution control may be positively correlated with increases in firms' revenues.

In order to summarize the methodological implications of the present study, it is useful to first remind ourselves of the nature of the data in applied production theory studies. In particular, the fact that a great majority of the empirical production theory studies are carried out using non-experimental data has important implications also for the approach used. Data are usually collected by central statistical offices or other authorities for various purposes, of which one may be research. As a result, primary data construction is rarely under the researcher's control. This lack of control and the fact that the data has not been collected for a particular research study often causes a lack of data precision and measurement errors. The stochastic approach allows measurement errors and random shocks to enter the model. Indeed, the particular characteristics of the data usually have an important influence on the specification of the econometric model, on the choice of the estimator, on the properties of the estimates and on inference. Moreover, the great advantage of the stochastic method is the possibility to use standard statistical tests to guide the model selection. On the other hand, the deterministic approach either assumes that there are no errors in the data or that they are one- sided. Moreover, the possibilities for "specification search" are very restricted for the deterministic approach.

The results of this study confirm the above problems for the deterministic approach. For the data used in the present study, the results for the deterministic model were unstable and sensitive to small changes in specifications. Consequently, the methodological implication of the present study is that sensitivity analysis should play a much more important role in deterministic parametric linear programming models than has been the case so far. At the least, a number of different model specifications should be tried and the sensitiveness of the results to small changes in the data should be examined.

Footnotes

1. Distance functions have also been used in consumer theory (e.g. Malmquist 1953, Deaton 1979, Cornes 1992), but their applications have not yet gained as wide popularity as in production theory.

2. It may noted that Charnes et al. (1988) and Evans and Heckman (1988) have compared the parametric linear programming and econometric approaches in estimating a translog cost function.

3. Also, as is well known, Monte Carlo simulation is sensitive to the design of the experiment (e.g. Davidson & MacKinnon 1993). One objection to the use of Monte Carlo evidence is that the sample design can make the differences between the techniques one is comparing more similar or extreme than they might be with data from an actual survey.

4. Outputs are called *weakly disposable* if $y \in P(x)$ and $\theta \in [0,1]$ and $\theta y \in P(x)$; and *strongly disposable* if $v \le y \in P(x)$ then $v \in P(x)$. According to the weak disposability of outputs assumption it is possible to reduce one output at least in a way that the other outputs are reduced in the same proportion, with inputs held constant. For example, it is at least possible to reduce water pollution (output) by one-third, with a simultaneous decrease of pulp (output) by one third. However, it is also possible to reduce water pollution by one-third and pulp output by two thirds, or vice versa. Finally, and most importantly, weak disposability does not rule out the possibility that undesirable output is reduced simultaneously with an increase in desirable output. Thus, also positive shadow prices for pollution are consistent with the weak disposability assumption.

5. See Färe et.al. (1989) for a nonparametric distance function approach which computes shadow prices for pollution.

6. The parametric linear programming models were computed using the GAMS 2.25 program, and the stochastic models using the LIMDEP 6.0 and GAUSS 3.2 programs.

7. The random coefficients estimation method and the test for the model (under the null hypothesis of parameter constancy) has been suggested by Swamy, P. (1971), *Statistical Inference in Random Coefficients Regression Models*, Springer-Verlag. The test statistic is algebraically the same as the standard *F* statistic for testing (see Greene 1993): $H_0: \beta_1 = \beta_2 = \ldots = \beta_n$.

References

Aigner, D. and S.F. Chu (1968). On Estimating the Industry Production Function, *American Economic Review*; 826-839.

Althin, R. (1994). Shadow Pricing of Labor, University of Lund, Department of Economics Working Paper Series, No. 33/94.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982a). Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers, *Economic Journal* 92; 73-86.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982b). The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity, *Econometrica* 50; 1393-1414.

Chambers, R., R. Färe and S. Grosskopf (1994). Efficiency, Quantity Indexes, and Productivity Indexes: A Synthesis, *Bulletin of Economic Research*, 46; 1-21.

Charnes, A., W.W. Cooper, and T. Sueyoshi (1988). A Goal Programming/Constrained Regression Review of the Bell System Breakup, *Management Science*; 1-26.

Chavas, J-P. and T.L. Cox (1994), A Primal-Dual Approach to Nonparametric Productivity Analysis: The Case of U.S. Agriculture, *Journal of Productivity Analysis*; 359-373.

Clement, J., S. Grosskopf, V. Valdmanis. (1994). A Comparison of Shadow Prices and Reimbursement Rates of Hospital Services, unpublished mimeo, July 1994.

Coggins, J. S. and J. R. Swinton (1994). The Price of Pollution: A Dual Approach to Valuing SO2 Allowances, unpublished mimeo, University of Wisconsin, Madison, June 1994.

Cornes, R. (1992). Duality and Modern Economics, *Cambridge University Press.*

Davidson, R. & J.G. MacKinnon (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

Deaton, A. S. (1979). The Distance Function and Consumer Behavior with Applications to Index Numbers and Optimal Taxation, *Review of Economic Studies* 46; 391-405.

Evans, D.S. and J.J. Heckman (1988). Rejoinder: Natural Monopoly and the Bell System: Response to Charnes, Cooper and Sueyoshi, *Management Science*; 27-38.

Fried, H.O., C.A.K. Lovell & S.S. Schmidt. (eds.) (1993). *The Measurement of Productive Efficiency*, Oxford University Press.

Färe, R., S. Grosskopf, C.A.K. Lovell and C. Pasurka (1989). Multilateral Productivity Comparisons when Some Outputs are Undesirable: A Nonparametric Approach. *Review of Economics and Statistics*, February; 90-98.

Färe, R., S. Grosskopf, and J. Nelson (1990). On Price Efficiency, *International Economic Review*; 709-20.

Färe, R., S. Grosskopf, C.A.K. Lovell and S. Yaisawarng (1993). Derivation of Virtual Prices for Undesirable Outputs: A Distance Function Approach, *Review of Economics and Statistics*, May; 374-380.

Färe, R. and S. Grosskopf (1994a), *Cost and Revenue Constrained Production*, Springer-Verlag, Bilkent University Lecture Series.

Färe, R., S. Grosskopf, C.A.K. Lovell (1994b). *Production Frontiers*, Cambridge University Press.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994c). Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries, *American Economic Review*, March; 66-83.

Färe, R. and D. Primont (1995), *Multi-Output Production and Duality: Theory and Applications*, Kluwer Academic Publishers.

Greene, W.H. (1993 a), The Econometric Approach to Efficiency Analysis, in Fried, H.O., C.A.K. Lovell & S.S. Schmidt (eds.).

Greene, W.H. (1993 b), *Econometric Analysis*, Macmillan, second edition.

Grosskopf, S. and K. Hayes (1993). Local Public Sector Bureaucrats and Their Input Choices, *Journal of Urban Economics*, 33, 151-166.

Hetemäki, L. (1994a). The Impact of Pollution Control on a Firm's Production Technology and Efficiency: A Stochastic Distance Function Approach, in Environmental Economics (eds. Brännlund, R., B. Kriström & K.-G. Löfgren), Sweden, June 10-13, 1993. Swedish University of Agricultural Sciences, Department of Forest Economics, Report no. 106.

Hetemäki, L. (1994b). Do Environmental Regulations Lead Firms into Trouble? Evidence from a Two-Stage Distance Function Model with Panel Data, Finnish Forest Research Institute, Research Papers no. 509.

Lovell, C.A.K., S. Richardson, P. Travers and L.L. Wood (1990). Resources and Functionings: A New View of Inequality in Australia, Department of Economics, University of North Carolina, Working Paper Series, no. 90-8.

Lovell, C.A.K. (1993). Production Frontiers and Productive Efficiency, in Fried, H.O., C.A.K. Lovell & S.S. Schmidt (eds.).

Malmquist, S. (1953). Index Numbers and Indifference Surfaces, *Trabajos de Estatistica* 4: 209-242.

Nishimizu, M., and J.M. Page (1982), Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, 1965-78, *Economic Journal* 92(368); 920-936.

Oates, W., K. Palmer and P. R. Portney (1993). Environmental Regulation and International Competitiveness: Thinking About the Porter Hypothesis, Resources for Future, Discussion Paper 94-02.

Porter, M. (1990). *The Competitive Advantage of Nations,* New York: The Free Press.

Shephard, R.W. (1953). *Cost and Production Functions*, Princeton University Press.

Shephard, R.W. (1970). *Theory of Cost and Production Functions*, Princeton University Press.

Appendix I: Data and variables

The data was collected from two different sources; Industrial Statistics collected by the Central Statistical Office of Finland (Teollisuuden Yleislomake ja Energialomake) and water pollution statistics collected by the National Board of Waters and Environment (Vesiensuojelun A ja B lomake). Both type of statistics are based on annual questionnaires sent to all plants. Because the questionnaires sent by two different authorities for collection of different information were not necessarily coherent, some of the figures were checked and corrected by directly contacting the plants involved. Since the data is confidential and its collection requires permission from each of the firms, code numbers are used for the plants in order to make them unidentifiable.

Output. The pulp and paper output series include information on the value and quantity (tons) for sulphate pulp. The implicit price index for output is derived by dividing the value of output by the quantity of output.

Water Pollution. The principal aim of water pollution monitoring is to assess the waste water ingredients, their quantity and toxicity, to control the compliance with permit conditions and to assess treatment efficiency and factors affecting efficiency. The monitoring is done according to a program approved by the supervising authority, i.e., the local water authority. The monitoring is carried out by both the official water laboratory (of which there are around 20 in the whole country) and by the pulp plants themselves. The local water authority gives limiting values at the plant level for the discharge in terms of total load per time unit or specific load per ton of product. In general, the limits must be attained as mean values for 1,3, or 6 months, depending on the size and type of plant. The water pollution statistics concerning the quantities of effluents of the pulp plants is considered to be of good quality by the National Board of Waters and Environment.

The water pollution statistics used in the present study consists of information on the flow of waste water (m3/a), biological oxygen demand (BOD7) (t/a) and suspended solids (t/a).

Labor. The information on labour input consists of data on both production and non -production (white collar) workers total numbers, hours worked, and wages and social security costs. Social security costs are not available for 1972-73 and so were estimated using the procedure outlined in Mäisti (Tulonjako paperiteollisuudessa vuosina 1955-1977, Työväen Taloudellinen Tutkimuslaitos, tutkimusselosteita 8:1979). The quantity of labour input is measured as the hours worked. Since there may be differences between production and non-production workers that is not reflected in the number of hours worked, the Divisia (or discrete time Törnqvist) index was used to compute an aggregate index of hours worked.

Capital. As is well known, the construction of data series for capital stock and price (user cost) of capital poses fundamental difficulties. For a clear exposition of these issues, see, e.g., Berndt, E. (1991). *The Practice of Econometrics*, Addison-Wesley.

The capital series consists of information on annual (1974-1990) purchases of capital goods (a), basic improvement costs (b), sales (c), and rented capital goods (d) of 6 different classes of capital assets (1. residential buildings, 2. non-residential buildings, 3. machinery, instruments and tools, 4. transportation equipment, 5. land and water structures, 6. other material investments). The gross investment series (e) is constructed as e = a + b + d - c. From 1972-73 there are also data on the fire insurance values of the different classes of capital assets. The 6 different classes of capital assets were first aggregated into two groups, namely, buildings = 1 + 2 + 5 and equipment and machinery = 3 + 4 + 6.

The replacement cost values of fixed capital assets were calculated from the perpetual inventory formula, $K_t = (1-\delta_t)K_{t-1} + I_{t-1}$,where K_t is the capital stock at the beginning of time t, δ is the constant rate of depreciation, and I_{t-1} is investment in period t-1. In order to obtain the starting (or benchmark) values for the capital stock we assumed equality of fire insurance cost and historic cost valuations of the capital stock in the first year of the data (1972) (Nickell et. al. 1992) have noted that "the choice of an accurate benchmark may be largely irrelevant" in a fixed effects panel data model).

In order to calculate the constant exponential rate of depreciation, the procedure given in Kuh, E. and R. Schmalense (*An Introduction to Applied Macroeconomics*, North-Holland, 1973) was used. According to this procedure the depreciation rate is calculated using the equation, $(1-\delta)L = X$, where L is the average service lives of capital assets and X is the value of capital assets as a percentage of their initial values at the end of their average service lives. It was assumed that, of the initial value of equipment and machinery, 10 percent is left after 32 years in the paper industry and after 25 years in the pulp industry. The corresponding figure for buildings was assumed to be 65 years for both industries. These figures for the service lives of capital assets are higher than those reported in the National Accounts. The figures used here are based on Simula (Tuottavuus Suomen metsäteollisuudessa. Licentiate thesis, University of Helsinki, Department of Social Economics of Forestry, 1979) rather than the more simple calculations of the Central Statistical Office. However, the figures should still be regarded as crude approximations. The above assumptions imply values of δ of 8.8% for equipment and machinery and 3.5% for buildings (For comparison, e.g., Nickell et. al. 1992 use the values 8.19% and 2.5%, respectively for the UK manufacturing industry). Finally, the replacement cost valuation of *total* fixed capital assets is calculated as the simple sum of the fixed capital assets of plant and machinery and buildings.

Materials. The data on intermediate materials consists of information on the value of materials. This is a "catch-all" variable which includes data on various inputs with different units. The important problem in constructing a materials input variable is that, as usual, there is no data on the quantity or price of this "input". Since it is essential to determine how much change in value can be considered a result of changes in quantity over time and across plants rather than in the prices, relevant price indexes or deflators must be found. In the present study the production price index for manufacture of paper and paper products was used as a deflator.

Appendix II: Model Diagnostics*

Autocorrelation. In order to check the residual autocorrelation, the autocorrelation functions and *Ljung-Box* Q-statistic were computed. The Q statistic tests the hypothesis that all of the autocorrelations are zero (three lags in the present case). The residual analysis was carried out for each plant separately. The results showed that the null hypothesis of no autocorrelation (at the 5% sig. level) was rejected for 5 plants out of the total of 8 plants in the Cobb-Douglas model, for 3 plants in the Cobb-Douglas fixed effects model, and only for one plant in the restricted translog fixed effects model. Thus, it appears that the likely cause for the autocorrelation problems in Cobb-Douglas specifications is the unsatisfactory functional form specification, since autocorrelation is not a serious problem in the restricted translog form.

Heteroskedasticity. The heteroskedasticity-consistent standard errors were used. However, in turned out that they do not differ significantly from the "raw" standard errors. Also, the White test indicated that the residuals are homoskedastic.

Normality. The *Shapiro-Wilk W-test* of normality was computed for the whole sample and for residuals from each of the model specifications. The null hypothesis of normality could be accepted at the 5% significance level for the whole sample and for all the model specifications. The results from testing for normality for each plant separately showed that normality could be accepted in the TRFE and CDFE models for all of them, except for two plants, at the 5 % significance level. However, when 2 outlier observations for plant 5 and one outlier for plant 7 were removed, the residuals were normally distributed. The null of normality was rejected in half of the cases for the Cobb-Douglas specification.

Orthogonality. The results in *Table 5.2* maintain the assumption of orthogonality between the error term and the regressors. In order to check whether the potential correlation of the right-hand side variables with the error term is great enough for the results of model 5 to be biased, the *Pearson correlation coefficients (r)* and the significance level (or probability level, *p*) of the respective correlations were computed and the *Hausman specification test* was run. In addition the relationship between the residuals and the right-hand side variables was examined by looking at the slopes of the regression lines from regressing the residuals on each of the exogeneous variable in turn. Variables that are not correlated are not necessarily independent, except for the case of the joint normal distribution, in which a lack of correlation does imply independence. The correlation coefficients showed that the correlation is rather low, the maximum value being 0.12, and none of the correlation coefficients was significant at the 1% level.

The Hausman specification test was consistent with the above results (Greene 1993). The test indicated that OLS is an efficient estimation method, in contrast to the instrumental variables estimator.

Multicollinearity. Multicollinearity led to rejection of the full translog specification. Due to the singularity of the Hessian matrix, this specification could not be estimated. The severity of multicollinearity in the remaining model specifications was examined using various methods (e.g., *stepwise regression*, *redundant variable tests* and *ridge regression*). These analyses showed that the shadow price results are not very sensitive to the possible remaining multicollinearity.

Testing parameter restrictions. The parameter restriction tests (*F-test and Chi2-test*) were carried out to examine which of the functional forms should be preferred and to see whether the homogeneity restriction is valid. The test results rejected the Cobb-Douglas specification in favour of the restricted translog specification. The homogeneity restriction consists of four different parts since there are three outputs, i.e., one part imposes $\sum_{m=1}^{M}\gamma_m = 1$, and three parts set $\sum_{m'=1}^{M} \gamma_{mm'} = 0$. The first restriction is the maintained hypothesis, since it also transforms the dependent variable and imposes the weak disposability of outputs. On the other hand, the latter three restrictions were rejected on the basis of the F-test at the 5% significance level but not at the 10% significance level. The estimation results are based on a specification where the three latter restrictions are set. Although the statistical test indicates problems with this restriction, if one were to relax it, the remaining function would not be fully consistent with a well-defined output distance function. Moreover, the results from a model in which the homogeneity restriction was imposed in one part, i.e.

$$\gamma_{BODFL} + \gamma_{BODQ} + \gamma_{BOD2} + \gamma_{FLQ} + \gamma_{FL2} + \gamma_{Q2} = 0,$$

showed that the shadow prices are not very sensitive to the specification of the restriction (the restriction was accepted by the F- and Chi2 tests).

Functional form specification. The RESET test due to Ramsey was used to test the null of correct specification of the original model against the alternative that the squared fitted values have been ommited. The coefficient for the squared fitted term was not significant, thus giving support to the restricted translog form.

**The detailed results of the plant specific diagnostics tests are available from the author on a request.*

Table 4. Hypothesis tests for restricted translog fixed effects model

1. H_0: Cobb-Douglas technology

   $F_{6,114} = 2.61$ (reject at the 5% level)

2. H_0: Homogeneity of degree +1 in outputs

   $F_{3,114} = 3.40$ (reject at the 5% level; accept at the 10% level)

3. H_0: Outputs are exogeneous

   Hausman test $\chi^2_3 = 0.31$ (accept at the 5% level)

4. H_0: No autocorrelation in residuals

   Box-Ljung Q-statistics (accept at the 5% level, see Appendix)*

5. H_0: Residuals normally distributed

   Shapiro-Wilkins W-test = 0.98 (accept at the 5 % level, see Appendix)*

6. H_0: Residuals are homoscedastic

   White -test $\chi^2_{48} = 55.7$ (accept at the 5% level)

7. H_0: Functional form correctly specified

   RESET test, $F_{1,112} = 1.36$ (accept at the 5% level)

Table 5. Panel model specification tests

Test Statistics for Panel Data Model Selection (see Greene 1993 a)

Model	Log-likelihood	Sum of Sqr Resid	R^2
(i) Constant term only	-154.2	67.7	0
(ii) Group dummies only	-91.0	29.5	0.56
(iii) Pooled model without plant or period effects	86.5	2.8	0.96
(iv) Fixed Effects	105.7	2.2	0.97
(v) Two Factor Fixed Effects	178.2	0.8	0.98
(vi) Two Factor Random Effects		5.1	0.92

Lagrange multiplier -test: model (vi) vs. model (iii) $LM_2 = 77.9$

Hausman test: model (v) vs. model (vi) $\chi^2_{12} = 14.8$

Random coefficients model vs. model (iii) $\chi^2_{91} = 364.7$ ("Swamy test")

ISBN 951-40-1422-7
ISSN 0358-4283