FINNISH FOREST RESEARCH INSTITUTE, RESEARCH PAPERS 609, 1996

Essays on the Impact of Pollution Control on a Firm: A Distance Function Approach

Lauri Hetemäki

HELSINKI RESEARCH CENTRE

METLA

METSÄNTUTKIMUSLAITOKSEN TIEDONANTOJA 609, 1996
FINNISH FOREST RESEARCH INSTITUTE, RESEARCH PAPERS 609, 1996

Essays on the Impact of Pollution Control on a Firm: A Distance Function Approach

Lauri Hetemäki

Academic dissertation to be presented, by the permission of the Faculty of Social Sciences of the University of Helsinki, for public examination, at Metsätalo, lecture room 2, Unioninkatu 40 B, on October 18, 1996, at 12 a.m.

HELSINGIN TUTKIMUSKESKUS – HELSINKI RESEARCH CENTRE
Helsinki 1996

Hetemäki, L. 1996. Essays on the Impact of Pollution Control on a Firm: A Distance Function Approach. Finnish Forest Research Institute, Research Papers 609, Metsäntutkimuslaitoksen tiedonantoja 609 ISBN 951-40-1525-8, ISSN 0358-4283.

The study analyzes the impact of pollution control and environmental regulations on the production of Finnish sulfate pulp plants. The study consists of three separate essays on the subject and a review chapter, which surveys the literature on empirical applications of distance functions. In each essay the output distance function is used as an analytical tool to model the multi-output production technology of pulp plants. The empirical models are estimated using plant level panel data from 1972-90. The study contributes to the existing literature by comparing the different methods of estimating distance functions, by extending the current estimation methods, and by providing a framework to analyze empirically the so-called *Porter hypothesis*. The results indicate that the reduction of the total amount of waste water effluents (unregulated bad) have increased the productivity and thus revenues of the plants. On the other hand, the impact of the abatement of biological oxygen demand (BOD) (regulated bad) is more ambiguous. Depending on the model specification the shadow price for BOD is either positive or negative. Moreover, in most of the models the estimated BOD parameter is not statistically significant at the conventional significance levels. The results also show that an increase in the intensity of environmental regulation (standards) has led to a decrease in the production efficiency of the pulp plants. Consequently, the results may be interpreted as not supporting one of the central premises of the Porter hypothesis. From the policy point of view, the results seem to suggest that environmental regulation should have been targeted at pollution-prevention measures (closing the water system), rather than at end-of-pipe measures (e.g. biological treatment lagoons).

Key words: pollution control, distance function, pulp industry, Porter hypothesis

Author's address: Finnish Forest Research Institute,
Unioninkatu 40 A, 00170 Helsinki, Finland.
Electronic mail: Lauri.Hetemaki@metla.fi

Publisher: Finnish Forest Research Institute; Project HE 3013.



Recyclable product with low emissions during production
Hakapaino Oy, Helsinki 1996

Acknowledgments

I would like to thank my examiners Pekka Ilmakunnas and Erkki Koskela for their valuable comments and helpful criticism. Pekka provided substantial help and advice from the start of the thesis work. Erkki has also been an important link at the Department of Economics, organizing and encouraging my work from beginning to end.

I am grateful to Jari Kuuluvainen for sightful comments at various stages of the present work. Moreover, he has acted as my project leader and collaborator, and he has been a friend with whom it has been a great pleasure to work. Matti Palo has provided energetic support for my thesis work, both by helping to organize the financing and by providing much encouragement.

The original idea of working with the distance functions came from Robert Chambers, who introduced me to the subject in a course in Denmark. Through him I got acquainted with the work of Rolf Färe, Shawna Grosskopf and Knox Lovell. They all sent me still unpublished papers, which helped me to get in touch with up-to-date literature on distance functions right at the start. Moreover, in occasional meetings at conferences, and through letters and e-mails, I have received many useful comments and relevant criticism from them, which have helped me to correct at least some of the mistakes in my work. I would also like to thank Tim Coelli for providing me with his program *FRONTIER 4.1* and his valuable comments on Essay III.

Mika Hakuni worked as a research assistant, helping with matters related to the data and with computation of the linear programming models. He not only did this work effectively, but provided sharp comments on various parts of the present study. My colleagues at the Nordic Council of Ministers project (*"NERP"*), Runar Brännlund, Bengt Kriström and Eirik Romstad, and Olli Tahvonen at the Academy of Finland project (*"Environmnet, Economy and Welfare"*), provided an interesting and productive research environment for the study. I am grateful for having had the opportunity to work with them. Runar and Olli have also provided valuable comments on parts of the thesis. Colleagues at the Finnish Forest Research Institute - too many to mention them all - have helped and supported my work on the thesis in various ways, for which I like to express my sincere gratitude.

Acknowledgments are also extended to Elina Karhu, Maija Kallio-Mannila, and Veli Suihkonen, who helped with different matters related to the collecting, processing and interpretation of the data, to Pertti Ruuska for drawing the figures, to Glenn Harma for checking the language, and to Pirjo-Riitta Lind for proof-reading.

I am indebted to Finnish and Nordic taxpayers who, willingly or unwillingly, have financed the study through the Academy of Finland, the Finnish Forest Research Institute, and the Nordic Council of Ministers.

My deepest gratitude goes to Eeva for her patience and also for diverting my attention to other important aspects of life. At last but not least, the long walks in the woods with Saana have been necessary for recharging my batteries.

Helsinki, September 1996 *Lauri Hetemäki*

CONTENTS

1. Background and Aim and Scope of the Study 7

1.1 Background 7
1.2 Pollution Control and the Firm 8
1.3 Aim and Scope of the Study 10

2. The Impact of Pollution Control on a Firm: A Distance Function Approach 13

2.1 The Conventional Approach 13
2.1.1 Theoretical Background 13
2.1.2 The Conventional Empirical Studies 15
2.2 The Distance Function Approach 20
2.2.1 Background 20
2.2.2 The Output Distance Function 22
2.2.3 The Derivation of the Shadow Prices of Bads 25
2.2.4 The Empirical Applications of Distance Functions 30
2.2.5 Nonparametric Linear Programming Studies 30
2.2.6 Parametric Linear Programming Studies 32
2.2.7 Econometric Studies 33
2.2.8 Concluding Remarks 41

3. Summary of the Essays 43

Notes 50
References 52

ESSAYS:

I. **The Impact of Pollution Control on the Pulp Industry: A Comparison of Deterministic and Stochastic Approaches** 57

II. **Estimating Shadow Prices for Bads Using a Two-Stage Stochastic Distance Function Model** 87

III. **Do Environmental Regulations Increase Production Efficiency? Evidence from the Pulp Industry** 119

Appendix I	Institutional Setting for Pulp Production and Effluent Control	150
Appendix II	Data and Variables	157
Appendix III	Panel Data Models	161
Appendix IV	Test Statistic	164

1. Background and Aim and Scope of the Study

1.1 Background

The firm and its production process has been one of the central concerns of the environmental economics literature since the subject area started to emerge as a separate branch of economic literature in the 1960s. The roots of the environmental economics literature are often traced back to the externality theories of Marshall and Pigou and the public goods theories of Wicksell. However, it was not really until the late 1960s and 1970s that the issues of production externalities of firms were addressed explicitly in the context of pollution (although Pigou already in 1932 used as an example of externality the impact of factory smoke on laundry services). The main interest of the early literature was, and still largely is, in what type of policy instruments should be used to optimally control the pollution discharges of firms. Typically, the central theme of this strand of literature has been the question of how to design the optimal standard, tax, subsidy, or other control measure to reduce firms' discharges to a fixed target level. Moreover, the other typical feature of the early literature was that it was predominantly theoretical. Indeed, the debate remained largely on the theoretical level until the late 1970s.

There now exists a large theoretical environmental economics literature on questions such as: What is the optimal amount of pollution or pollution control?; What policy instruments are most (cost) efficient in achieving it?; and What are the pollution control costs to firms? (see e.g. Baumol and Oates 1988, Cropper and Oates 1992, and Tietenberg 1992). In the present study, the analysis is restricted to the last question. Pollution control costs are the costs incurred in changing an economic activity in some respect in order to reduce discharges. For example, a polluting firm might invest in external treatment facilities, in internal process changes, in changing its product or input mix, in relocating, or in pursuing some combination of these measures, or it might even close down. Whatever it chooses to do, the actions will most likely show up as changes in production and/or operation costs.

Interest in conducting empirical studies on the impact of environmental regulations on producer behavior was stimulated to great extent by a number of studies in the late 1970s which started to analyze the factors behind the productivity slowdown in Western economies. After the rapid economic growth of the 1950s and 1960s, there was a

general tendency toward slower economic (and productivity) growth in the industrialized countries. One of the important factors that was suggested as a possible cause for this slowdown was environmental regulation. In particular, it was argued that abatement investment "crowds out" alternative capital investment in plant and equipment. As a result, a number of studies tried explicitly to analyze to what extent this was indeed the case (e.g. Denison 1978, The Data Resources (DRI) 1979, Meyers and Nakamura 1980, Gollop and Roberts 1983, Conrad and Morrison 1989, Jorgenson and Wilcoxen 1990). Besides their contribution to the productivity growth literature, these studies served at least two other important functions. First, they begun to emphasize the importance of analyzing the impact of environmental regulation on a firm, industry or economy, besides the more traditional concern for the optimal control policy to attain a target level of abatement. Secondly, these studies gave impetus to empirical studies in the environmental economics literature.

1.2 Pollution Control and the Firm

In the environmental economics literature, the terms pollution control and environmental regulation are often used synonymously.[1] Although, in practice it may often be difficult to separate the impacts of these two factors on production, it is important to be aware of the difference between the two concepts. This is because of the following reasons. First, firms may control (reduce) pollution independently of regulation. For example, pollution may often be related to material waste and inefficiency of the production process, and consequently the firm has an interest in reducing pollution even if it is not regulated. Also, the incentive for firms to control pollution may originate from consumer preferences such as "green values" - again, independently of the regulation. In the economics and management literature there is ample evidence of both of these two factors (e.g., Kneese and Bower 1968, Myers and Nakamura 1980, Porter and van der Linde 1995). In summary, environmental regulation may be but *one* of a number of reasons why firms control pollution.

The basic questions concerning the effects of regulation are whether the regulation achieves its objective (e.g. a certain standard of water quality) and what costs are associated with the regulation. The present study is restricted to the latter issue. A further restriction on the scope of the present analysis concerns the extent to which the effects

of regulations and pollution control are considered. The initial incidence of much of pulp industry pollution control falls on the industry itself. The ultimate incidence depends on the ability of firms to shift the cost burden to consumers by raising prices or to workers by reducing employment and wages or to other factors of production. This study focuses on the initial incidence and is therefore restricted to a partial equilibrium analysis of how a firm (or industry) reacts to a change in its cost structure.

Whether firms control pollution because of environmental regulation or for some other reason, the introduction of a pollution abatement process will involve changes in the production process. Jorgenson and Wilcoxen (1990) identify three different response categories of environmental regulations. First, the firm may substitute less polluting inputs for more polluting ones. Second, the firm may change the production process to reduce emissions. Third, the firm may invest in pollution abatement devices. Naturally, these various measures may be adapted simultaneously. In the literature, the two first measures are known as *"pollution-prevention"* methods, and the third as an *"end-of-pipe"* measure.

Depending on the ease of substitutability of inputs, switching to cleaner inputs may be the least disruptive of the above three possible responses of the firm, since it does not necessarily require as extensive a reorganization of the production process as do the second and third responses. A high degree of substitutability between inputs implies low costs of environmental regulation, and vice versa.

The second response to pollution control is the process change, which involves the redesign of production methods to reduce emissions. Such internal process changes may have either a positive or negative effect on production of the "good" output (c.f. Kneese and Bower 1968, Barbera and McConnell 1990). For example, the internal process change may require more inputs for a given level of good output, thus having a negative impact on productivity. On the other hand, it may very well also reduce the requirement of inputs, due to a more efficient production process, and thus have a positive impact on productivity. Moreover, it has also been suggested that increasingly stringent environmental regulation may cause more capital turnover and hence modernization, so that the net effect may be increased productivity growth (c.f. Meyers and Nakamura 1980, Porter and van der Linde 1995).

The third response to pollution control is to invest in abatement technology, i.e., in the use of special devices to treat wastes after they have been generated. According to Jorgenson and Wilcoxen (1990), end-of-pipe abatement is often the choice for existing

firms that have to meet newly imposed standards. This type of investment in external treatment imposes a direct cost on the industry and thus raises the total input costs for a given level of output. The *net impact* of environmental regulations or pollution control on firms production performance depends on which of the above effects dominates and on what the returns are from being able to satisfy consumers' preferences, such as "green values" (In the present study the impact of consumer preferences on production is not analyzed. Indeed, during the study period 1972-90 green marketing strategies were still of only minor importance for pulp plant operations; they have become more prevalent only recently). Thus in the end, it is left to empirical analysis to establish the magnitude of the net effect in each particular case.

1.3 Aim and Scope of the Study

The three essays of the study aim to provide quantitative estimates of the impact of pollution control and environmental regulation on the Finnish pulp industry during 1972-90. In particular, two different aspects of these impacts are analyzed. First, the aim has been to establish what have been the costs of water pollution abatement for the pulp firms, i.e. to derive the "shadow prices" of water pollution control measures. To put it simply, these shadow prices provide an answer to the question: How much does it cost pulp firms to reduce water pollution in terms of foregone pulp output or revenues?

Secondly, the study analyses what has been the impact of environmental regulation (standards) on the production efficiency of the pulp mills. This question is related to the so called *Porter hypothesis* (Porter and van der Linde 1995, Oates et al. 1993). Since the Porter hypothesis has not been expressed formally there is some ambiguity as to exactly what the hypothesis is (Jaffe and Palmer 1996). However, from the literature it is clear that one of the central arguments of the hypothesis states that properly designed environmental standards can trigger innovation and production efficiency gains that may lead to absolute advantages over non-regulated firms (e.g. Porter and van der Linde 1995, Jaffe et al. 1995, Jaffe and Palmer 1996). This argument or hypothesis has recently been the subject of active debate in the environmental economics and management literature. Yet, as Jaffe and Palmer (1996) have noted, it appears that no study has empirically tested this hypothesis. The present study provides a framework for empirically analyzing this hypothesis. Moreover, the impact of *environmental regulation*

(water pollution standards) on the production efficiency of Finnish pulp plants is examined. Depending on whether this impact is negative or positive, the evidence can be interpreted as either rejecting or supporting a central premise of Porter's hypothesis vis-à-vis the pulp industry.

By restricting the analysis to the above two categories of impact of pollution control, we rule out an assessment of the benefits generated by a cleaner environment. However, this does not in any way mean that these benefits are unimportant. Indeed, the primary purpose of environmental regulation and pollution control is to increase the net benefits to society. But there are a number of ways a society can try to achieve a given level of pollution control, each resulting in different costs. A knowledge of these costs is important to enabling the implementation of the most efficient pollution control policy. For example, the more efficiently the control target is achieved, the more resources are left to other purposes, of which one may be to obtain further increases in pollution control. Moreover, the costs of pollution control can in principle be weighed against measures of the marginal benefits of pollution control, in order to guide the optimal level of regulation.

The theoretical framework of the study is based on production theory and in particular on a *distance function approach*. The basic ingredients of the theory were already formulated by Shephard (1953 and 1970). By now, there exists an abundance of results concerning distance functions and their properties. The main advantage of the distance function approach is that it is more general than the conventional approaches, i.e. production, cost, profit, and revenue functions (the theoretical results obtained for distance functions have been recently summarized by Färe and Primont 1995). Also, the computation of direct distance functions requires only quantity data. This is of particular importance in environmental economics applications, since price data related to environmental compliance costs are often not available or are unreliable.

Although the theoretical framework on which the distance functions are based has been known for a long time (four decades), it is only recently that their usefulness in empirical applications have come to be appreciated. In particularly, the work of Rolf Färe and his colleagues has been influential in popularizing the use of distance functions. For example, the study by Färe, Grosskopf, Lovell and Yaisawarng (1993) was apparently the first study to use a distance function to derive shadow prices of undesirable outputs (pollution). This study is based on the linear programming approach, as are most of the existing applications of distance functions. It appears that only eight econometric distance function studies have been published so far (Lovell et al. 1990, Grosskopf and

Hayes 1993, Hetemäki 1994 a,b, 1995 a,b, Coelli and Perelman 1996, and Grosskopf, Hayes, Taylor and Weber 1996; unpublished manuscripts include Brännlund 1996 and Atkinson, Färe, and Primont 1996). Probably the most important reason for the paucity of econometric applications is the fact that the stochastic estimation of distance functions is more involved than the computation of linear programming models or the estimation of production, cost, and profit functions. Indeed, the stochastic estimation of distance functions is still being developed.

The remainder of the study is organized as follows. Chapter 2 provides a very brief and selected literature review on how the distance function approach can be used to analyze the impact of pollution control on a firm. This chapter helps to put the three essays into perspective. First, the conventional approach to the topic is provided, after which the recent distance function approach is introduced. Since the contributions of the present study rest partly on the development of the methods for estimating distance functions, the different methods that have been used to estimate distance functions are discussed, with particular emphasis on the econometric approach. Chapter 2 ends with some concluding remarks. After that, the three essays are presented. Chapter 3 provides a summary of the essays.

2. The Impact of Pollution Control on a Firm: A Distance Function Approach

In the present chapter, the distance function approach to modeling the impact of pollution control on a firm is described. However, in order to put the distance function approach in perspective, we first present the stylized facts concerning the conventional approach and some of its empirical applications. After that, the background of the distance function approach is given. In particular, the output distance function, which is the main analytical tool used in the three essays of the study, is presented in some detail. Moreover, since the estimation of the shadow price of pollution control is one of the main purposes of the study, the derivation of the shadow price of bad (effluent) output and its interpretation is described in the present chapter. This saves us from repeating the presentation in detail in the following essays. In the final part of the chapter, the different methods used to compute the distance function, in particular, the stochastic or econometric method, are discussed. The rather detailed presentation of the empirical approaches is justified on two bases. First, the empirical applications, comparisons and development of the different estimation methods for distance function is a central theme of the essays of the study. Secondly, empirical applications of (parametric) distance functions are relatively new and it appears that there is still no coherent presentation or survey of the topic in the literature.

2.1 The Conventional Approach

2.1.1 Theoretical Background

Without going in to great detail, the stylized facts are described for the conventional neoclassical theoretical framework of the impact of pollution control on the firm. Typically, the technology of the polluting firm is modeled using one of the two basic approaches. Either the effluent is included in the production or profit function as an additional input or, alternatively, abatement capital is included as an additional input in a cost function. In order to illustrate these approaches, consider the following simple model.

In the model, a firm's production function (f) consists of one "desirable" or "good" output (Y_1), capital input (K), labor input (L), and effluent (possibly a vector, Z). Further,

assume that the authority imposes a tax q on the firm's effluent. The exogenously given prices of output, capital, and labor are denoted by P_Y, P_K, and P_L, respectively. The profit maximizing firm's profit is then

$$(1) \quad \pi = P_Y f(K,L,Z) - P_K - P_L - qZ .$$

Differentiating with respect to *K*, *L* and Z, the first order conditions of profit maximization are:

$$\begin{aligned} & P_Y f_K(K,L,Z) = P_K \\ (2) \quad & P_Y f_L(K,L,Z) = P_L \\ & P_Y f_Z(K,L,Z) = q. \end{aligned}$$

The first order conditions, together with the production function, determines the output, input use and effluent as a function of product and input prices. It may be noted, that in this framework the introduction of the effluent does not in any way change the conventional model of firm. As Uimonen (1992) has noted, *"If the effluent is interpreted as an input and the effluent tax as the price of that input, the results of the standard theory of the firm apply. The inclusion of effluents in the production function adds nothing new to the theory"* (p. 11). Moreover, had we introduced an effluent standard instead of an effluent tax, this conclusion would not change either.

Alternatively, in the literature it has been popular to model the impact of pollution control on firm production using a cost function and to introduce the environmental constraint in terms of abatement capital or expenditure (e.g. Barbera and McConnell 1990). Thus, effluents are treated as a function of abatement capital (*A*). The abatement capital is assumed to be exogenous to the firm in the short run. The firm's cost function (c) may then be defined as

$$(3) \quad C = c(Y_1, P_K, P_L, A) .$$

In equation (3), abatement requirements shift the cost function by the amount of abatment expenditure. The employment of abatement capital may also lead to changes in the combination of inputs used to produce output Y_1. This indirect impact may be negative or positive (see Barbera and McConnell 1990). However, abatement capital is

again treated just like any other fixed input in the cost function, and nothing essential changes compared to the conventional cost function of a firm.

Recently, there have been attempts to modify the conventional framework in order to allow for a more rich description of the production technology of the polluting firm. One interesting approach is presented in Uimonen (1992), in which the polluting firm's production process is described by a multi-product process, consisting of three production units which are linked together. The first production process is the unit generating the basic "good" or "desirable" product, the second process is a side-unit, which is the abatement plant for the residuals, and the third process is the side-unit which produces a marketable side-product. Without going in to detail, Uimonen (1992) demonstrates that this framework generates somewhat different comparative statics results than the conventional single-product framework. In his theoretical study, Uimonen uses the framework to analyze the impact of a subsidy on abatement capital, an effluent tax, and a standard applied to the polluting firm.

Although Uimonen (1992) demonstrates the advantages of the more rich framework (joint multi-output approach) on modeling a polluting firm's production technology and it derives new interesting results, the main drawback of the approach for the present purposes is that it is tedious to apply for empirical work and it imposes unnecessary restrictions on the technology. In particular, the approach requires estimating also the side-product process, for which the output price has to be non-zero. Also, it requires price data in general and assumes technical efficiency in production.

2.1.2 The Conventional Empirical Studies

The conventional neoclassical single-output models seem to indicate that the empirical assessment of the impact of pollution control and regulation policy on producer behavior would also be straightforward. However, in practice reliable data on abatement costs and the resulting reduction in effluents are often not available, at least at the firm or plant level. Furthermore, the problem has been how to derive each firm's marginal pollution control costs. Indeed, in the most common case, where there is no market for pollution, one does not observe (directly) the price of pollution. Thus, one usually has to try to infer the marginal control costs indirectly.[2] The empirical literature has proved to be innovative, and a variety of approaches have been used.

Some of the interesting early econometric applications include Muller (1978) and Sims (1979). Muller (1978) estimated a fixed-coefficients production model using annual time series data (1949-69) for the Canadian pulp and paper industry to simulate the reaction of the industry to those changes which might be expected from mandatory pollution controls. This is one of the earliest attempts to analyze the impact of pollution control on the pulp and paper industry using an econometric model. Muller found that the increases in average total production costs resulting from secondary treatment of all water-borne wastes varied from less than 3 per cent for sulfate pulp mills to around 9 per cent for sulfite pulp mills. Moreover, the results indicated that overall output in the industry would not decline by more than about 2 per cent if cost increases averaged about 6 per cent. Sims (1979) examines the impact of an effluent charge on waste water in the Canadian brewing industry, using a combination of time-series and cross-sectional data (26 observations from 3 breweries) to estimate a translog cost function. He found that a one per cent increase in the charge for biological oxygen demand (BOD) and suspended solids (SS) substances would lead to a 0.57 per cent reduction in BOD effluents and a 0.45 per cent decline in SS effluents, respectively.

Since the early 1980s, the number of empirical studies which have examined the impact of environmental regulations on economic performance have increased significantly (e.g. Myers and Nakamura 1980, Pittman 1981, 1983, Gollop and Roberts 1983, Conrad and Morrison 1989, Jorgenson and Wilcoxen 1990, Barbera and McConnell 1990, and Gray and Shadbegian 1993, 1995). These studies have examined various effects of pollution control policy, including the direct impact of regulation on firms due to the diversion of resources into abatement capital and an indirect effect due to changes in the production process and usage of inputs. The ultimate aim of these studies has often been to measure the effect of pollution regulations on *total factor productivity (TFP)* growth. On the other hand, some of these studies have also analyzed the cost effectiveness of the regulatory system itself. For the most part, these studies are based on industry level data or, in a few cases, on plant level cross-section data. There have been very few empirical studies which have attempted to analyze the effects of regulations using *plant level panel data*, one exception is Brännlund and Löfgren (1996).[3]

The most popular approach in the above studies to try to incorporate the impact of environmental regulations on firm, industry, or economy has been to use the total amount of pollution control expenditures as a measure of the impact of pollution control.

This is the approach used for example in the Barbera and McConnel (1990) study discussed above. The essence of the approach is that environmental regulations force firms to invest in abatement capital and this crowds out productive investments. However, Barbera and McConnel (1990) also take into account that abatement capital may have an "indirect effect" by making conventional capital obsolete more quickly and may increase investment in conventional capital and improve productivity. In order to quantify these impacts, Barbera and McConnel estimate a translog restricted cost function using annual time series data for five industry sectors. They include abatement capital as a input in the cost function in addition to the more traditional inputs (prices). In their framework, the direct abatement effect always lowers TFP growth. They find that in most of the industries, the indirect effect causes productivity to decrease, though in some industries productivity rises during certain periods. In all the industries, abatement requires more energy and conventional capital. The net impact of abatement investment is to lower productivity in all five industries, with abatement accounting for between 10 % and 30 % of the decline in productivity in the 1970s.

Conrad and Morrison (1989) examine the impact of pollution control on productivity in the U.S., Canada, and Germany. They construct a *nonparametric* model that explicitly recognizes the difference between pollution abatement capital and "productive" capital and then use a framework that purges the impact of abatement capital on total factor productivity growth. In the empirical application the bias resulting from abatement capital is computed for the aggregate manufacturing sectors of the U.S., Canada, and Germany. The principal finding is that the bias, and thereby the impact of increased pollution abatement capital on productivity growth, is modest.

Both of the above studies represent the typical approach of the majority of studies in the 1980s, in that these two studies are based on aggregate time series data and try to infer the impact of pollution control from abatement capital expenditures. Examples of a somewhat different approach are the plant level studies by Pittman (1981) and Gollop and Roberts (1983). Pittman (1981) examines the efficiency of the regulatory system by comparing shadow prices of water pollution effluents between different mills, and he analyzes whether it is less costly for the larger mills to control pollution than for the smaller ones. Pittman estimates a translog production function using cross section data from 30 integrated paper mills in Wisconsin and Michigan. According to the results, the systematic differences in the shadow prices of water pollution parameters (marginal treatment costs) indicate serious inefficiencies resulting from the prevailing regulatory

system. Secondly, Pittman interprets the positive correlation of pollution control intensity with economies of scale as evidence that pollution control may have negative allocational impacts on the industry. That is, treatment requirements increase the minimum efficient size of a plant, thus increasing barriers to entry and decreasing competition in the industry.

Gollop and Roberts (1983) study the impact of sulfur dioxide emission restriction on the rate of productivity growth for a sample of 56 electric power utilities in the U.S. over the period 1973-79. The empirical analysis is based on a translog cost function, which includes a firm-specific *regulatory intensity variable*. The idea is that an increase in the regulatory intensity shifts the cost function upwards and decreases productivity growth. Gollop and Roberts find that annual average productivity growth in the electric utility sector would have been 0.44 percent higher in the absence of the sulfur dioxide regulation.

Conclusions from the above studies, and from the studies summarized e.g. in Christainsen and Tietenberg (1985) point toward the idea that pollution control and environmental regulations have had an adverse effect on firm or industry performance but that it has not been large in magnitude. More recently, Jaffe et al. (1995) conclude in their survey of the impact of environmental regulations on the competitiveness of the U.S. manufacturing sector that *"there is relatively little evidence to support the hypothesis that environmental regulations have had a large adverse effect on competitiveness, however that elusive term is defined"* (p. 157). Finally, it may be noted that apparently only two studies (Uimonen 1986 and Hakuni 1994) have attempted to analyze the impact of environmental regulation or pollution control on Finnish industries.[4]

The above studies, which we have termed as representing the conventional approach, suffer from a number of weaknesses. First, as concerns the studies using aggregate data (i.e. most of them), they are subject to potentially serious aggregation errors. As was already indicated earlier, if pollution regulations are set e.g. at the plant or firm level, it is also important to use data at the same level. Besides the aggregation bias, the conventional studies suffered from the restrictions of the theoretical framework employed. Most of the studies published before the 1990s are based on production, cost or profit functions, with the pollution variable modeled *indirectly* using the abatement capital expenditures as a *factor input*. Thus, the production models have not analyzed the *joint* production of multiple outputs (good and bad outputs) and the pollution outputs

(emissions) have not been incorporated directly into the models of production technology (with the exception of Pittman 1983). However, as shown e.g. by Pittman (1983), Uimonen (1992), Färe et al. (1993) and the "materials balance" approach (Kneese et al. 1970), the joint output framework seems intuitively preferable. For example, in the pulp industry case, the fresh water entering as an input in the production process does not contain the regulated substances. It is only in the production process that the water gets contaminated with different kinds of substances and is transformed into an undesirable *output*. The importance of including bad outputs directly in models of producer behavior was stressed already two decades ago by Shephard (1974, p. 205), who noted that *"... , for the future where unwanted outputs of technology are not likely to be freely disposable, it is inadvisable to enforce free disposal of inputs and outputs. Since the production function is a technological statement, all outputs, whether economic goods are wanted or not, should be spanned by the output vector y."*

Moreover, the conventional studies have implicitly assumed that the firms are operating on the production frontier and that pollution control or regulation does not have an impact on production efficiency. However, as later studies have indicated, these assumptions are unlikely to hold in many instances (e.g. Färe et al. 1989, Hakuni 1994, Yaisawarng and Klein 1994, Porter and van der Linde 1995). Finally, with the exception of production functions, all the other conventional functions used to represent production technology require data on prices. This is an important shortcoming particularly for empirical environmental economics studies. As Jaffe et al. (1995) have recently stated, *"Even for the United States where data on environmental compliance costs are relatively good, compliance expenditure data are notoriously unreliable. The problem is more pronounced in other OECD countries, whose environmental agencies have not typically tracked environmental costs. Thus, we may have found little relationship between environmental regulations and competitiveness simply because the data are of poor quality"* (p. 158).

As will be argued below, the so-called distance function approach can potentially avoid the above problems. In the context of the impact of environmental regulations on firm or industry, the distance function approach and the weak disposability of undesirable outputs was apparently first suggested implicitly in the studies by Färe et al. (1983, 1986). They used nonparametric linear programming, the weak disposability assumption and the axiomatic approach of Shephard to compute the impact of environmental regulation on efficiency, although not explicitly in terms of distance functions but rather in

terms of *output correspondences*. The parametric linear programming study by Färe et al. (1993) was the first study to explicitly use the distance function approach to derive shadow prices of undesirable outputs, and Hetemäki (1994 a,b, 1995 a,b) first used an *econometric* distance function approach to this end.

2.2 The Distance Function Approach

2.2.1 Background

The concept of a distance function (also known as a gauge function, transformation function, or deflation function) comes from the mathematical theory of convex sets and was introduced into economics in the early 1950s in a different context by Debreu (1951), Malmquist (1953) and Shephard (1953). Debreau formulated the "*coefficient of resource utilization*", which offered a radial measure of productive efficiency. Malmquist, a Swedish statistician, introduced the distance function into consumer theory: he formulated a quantity (or standard-of-living) index that is dual to the "true" cost-of-living index. Specifically, "the *Malmquist quantity index*" is given by the ratio of values of the distance function (at different quantity situations, normalized on a given utility level). Shephard (1953) formulated the distance function in the production theory context. Shephard introduced the input distance function and the duality result between the input distance function and the cost function.

The profound impact of these contributions only became apparent some two decades after they were initially published. These studies were rediscovered in the 1970s when duality theory in production and consumer theory began to attract widespread interest. Shephard's application to production theory, especially as refined in his later volume (Shephard 1970), finally found its way into the important contributions of modern production theory, such as those of Fuss and McFadden (1978). Shephard (1970) provided the duality result between the output distance function and the revenue function. The indirect duality versions of the input distance-cost function and output distance-revenue function dualities were first formulated in Shephard (1974).

Shephard's Duality Results

TECHNOLOGY	DUALITY	VALUE FUNCTION
Input Distance Function	⟷	Cost Function
Output Distance Function	⟷	Revenue Function
Indirect Input Distance Function	⟷	Revenue Indirect Cost Function
Indirect Output Distance Function	⟷	Cost Indirect Revenue Function

Färe and Primont (1990) proved a duality theorem for the output distance function and the indirect output distance function, thus providing a link between direct and indirect models of production. In addition, Färe and Primont (1995) proved a duality theorem for the indirect cost function and the indirect revenue function which links the input-based and output-based models of production. A comprehensive exposition of all the duality results in production theory is presented in Färe and Primont (1995).

Distance functions are a tool for representing *production technology*, just as are production, cost, and profit functions, i.e. for showing "how inputs are turned to outputs." However, they have a number of desirable properties, which the more conventional functional representations lack. First, in comparison with the production function, the distance function allows one to model *multiple output* and *joint* production technologies. Second, distance functions provide a measure for technical efficiency and an index of resource utilization. This is of great theoretical and practical significance, for the distance function serves two important roles simultaneously. It provides a complete characterization of the structure of multi-input, multi-output production technology, and it provides a measure of the distance of each producer from the frontier technology. Thirdly, one important advantage of the distance function over the cost, profit and revenue functions (which also readily model multiple output technology) is that no maintained behavioral hypothesis (cost minimization or profit/revenue maximization) is required. The distance function only identifies the boundary or frontier technology and measures the distance to that boundary from every observation in the sample. Therefore, distance functions are particularly useful for modeling the behavior of e.g. public organizations (hospitals, schools, police force, etc.) and monopolies which do not necessarily follow conventional optimizing behavior. Indeed, distance functions allow explicit testing of hypothesis such as cost minimization and profit/revenue maximization.

Fourthly, the large number of duality results established for various distance functions allow one to derive shadow prices for outputs or inputs. This is important because often for reasons related to data availability, non-existence of markets for inputs or outputs, or the existence of imperfect competition, one does not know the relevant prices for inputs or outputs. Fifthly, the distance function can be computed merely with data on quantities of inputs and outputs; i.e. price data are not necessarily needed.

In summary, the distance function can be used for analyzing all the economic effects that are analyzed by applying conventional production, cost, profit or revenue functions. Moreover, because of its generality, the distance function is applicable even in cases where the latter functions cannot be used. In addition, Chambers, Färe and Grosskopf (1994, p. 1) have shown, that *"the distance function is the unifying notion which links efficiency measures, quantity indexes and productivity indexes."*

The above advantages come at some cost. The main disadvantage of distance functions is that their empirical application is not as straightforward as that of production, cost and profit functions. In particular, the stochastic (econometric) estimation of distance functions is more involved than that of the conventional functions (see below).

2.2.2 The Output Distance Function

Since all three Essays of the present study are based on the direct output distance function, we present the concept more formally below. It is also useful to have the theoretical background established before we turn to a discussion of the different empirical approaches to the computation of distance functions, later in the present chapter.

The conventional production function is defined as the maximal output that can be produced from an exogenously given input vector. The direct output distance function generalizes this notion to the multi-output case. It describes "how far" an output vector is from the boundary of the representative output set, given the fixed input vector. The output distance function can be defined in terms of the output set *P(x)*, i.e. the set of all output vectors that are technically feasible for input vector x. Suppose that a pulp plant employs a vector of inputs $x \in R_+^N$ to produce a vector of outputs $y \in R_+^M$, where R_+^N, R_+^M are non-negative N - and M -dimensional Euclidean spaces, respectively. The relationship between inputs and outputs is captured by the plant technology, which can

be described by the output set $P(x) \subseteq y$. The output set contains all output vectors that can be produced with the input vector *x*. It is assumed that the output set satisfies the maintained axioms of Färe (1988).[5] The formal definition of a direct output distance function in terms of the output set is

$$(7) \quad D_o(x,y) = \inf_{\theta}\{\theta > 0 : (y/\theta) \in P(x)\} \forall x \in R_+^N .$$

Equation *(7)* measures the largest radial expansion of the output vector *y*, for a given input vector *x*, that is consistent with *y* belonging to *P(x)*. The infimum (rather than minimum) is used because in the multi-output case it is possible that the minimum cannot be achieved (Färe and Primont 1995). However, for simplicity, we hereafter use the minimum. An intuitive interpretation of the output distance function is given with the aid of *Figure 1*. In order to determine the value of D_0 *(x, y)*, the output set *P(x)* is drawn in *Figure 1*. Let *y** be the intersection of the ray from the origin through *y* and the boundary of *P(x)*. The output distance function pushes output vector *y* as far from *0* as possible along the ray *0y**, whith *(y/θ)* remaining in *P(x)*. From *Figure 1* it is obvious that with $D_0 < 1$ we are inside *P(x)* (inefficient output), with $D_0 = 1$ on the boundary (efficient output), and with $D_0 > 1$ we are outside the producable output set *P(x)* (impossible output).

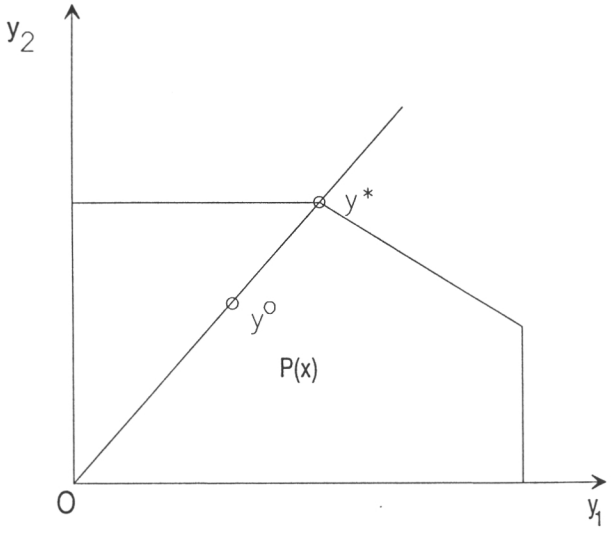


Figure 1. The output distance function

The output distance function can be used to measure the *Debreu-Farrell technical efficiency (DF)* (Debreu 1951, Farrell 1957). In terms of the above output set, the Debreu-Farrell measure can be defined as $DF(y,x) = \max\{\theta : \theta y \in P(x)\}$; and in terms of the output distance function $DF(y,x) = 1/D_o(y,x)$. Thus, the DF measure is the reciprocal of the value of the distance function and it gives the factor by which all outputs could be expanded proportionately if the production units were operating on the frontier. It is clear that $D_o(y,x) \leq 1$ and that the isoquant (efficient subset) is the set of y's for which $D_o(y,x) = 1$. For example, if $D_o = 1$, the firm can be regarded to be 100 percent efficient. For $D_o \leq 1$, the firm produces in the interior, and as this is less than the maximum possible production, it could be characterized as $100 * D_o$ percent efficient.

The output distance function will have, among others, the following properties (for a detailed description, see Färe 1988):

1. $D_o(0, y) = +\infty$ *for* $y \geq 0$ *, i.e. no free lunch.*
2. $D_o(0, x) = 0$ *for all* x *in* R_+^N *, i.e. inaction is possible.*
3. $x' \geq x$ *implies that* $D_o(x', y) \leq D_o(x,y)$*, i.e. the more input the less efficient.*
4. $D_o(x, \mu y) = \mu D_o(x, y)$ *for* $\mu > 0$*, i.e. positive linear homogeneity.*
5. $D_o(x,y)$ *is convex in y.*

Conventionally, the assumption of strong (or free) disposability of outputs is made. However, this assumption may exclude important production processes, such as undesirable outputs. For example, in the case of the pulp industry, the BOD and suspended solids (SS) effluents have been regulated and the plants cannot just "freely" dispose of these effluents. The difference between the strong and weak disposability assumptions is illustrated in *Figure 2.* In *Figure 2* the output set $P(x)^0$ is consistent with weak disposability of bad output, while the output *sets* $P(x)^1$ impose strong disposability of bad output. More formally, if the bad output, y_1, is weakly disposable, then $y_1 \in P(x)$ and $0 \leq \theta \leq 1 \Rightarrow \theta y_1 \in P(x)$. In *Figure 2* , the weakly disposable output set, $P(x)^0$, is described by the area 0ABCDE. This implies that the firm can reduce the bad output by a given amount (along the ray B0) by proportionately decreasing the good output. The intuitive explanation is that, with fixed inputs, the firm may have to remove some of its inputs from the production of good outputs to the abatement process, i.e. to the reduction of bads. It should be noted that all the output sets also assume that the good output is

null-joint with the bad, i.e. if a good output is produced in a positive amount, then some bad output must also be produced. If both outputs are *strongly* (or *freely*) *disposable*, the output set is defined by $P(x)^1$. More formally,

if $(y_1, y_2) \in P(x)$ and $0 \leq y_1^* \leq y_1, 0 \leq y_2^* \leq y_2 \Rightarrow (y_1^*, y_2^*) \in P(x)$.

This would imply that the output of a bad can be reduced without any reduction in good output (i.e. moving from point B to point A in $P(x)^1$) .

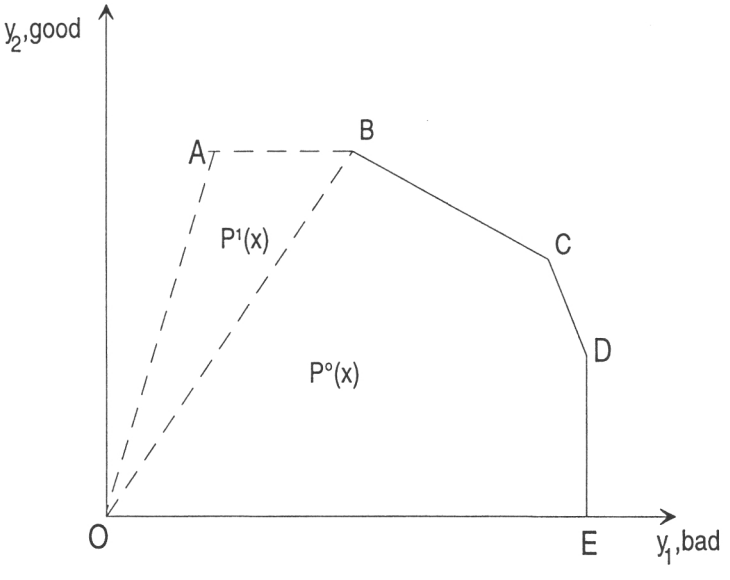


Figure 2. Disposability of Outputs.

2.2.3 The Derivation of the Shadow Prices of Bads

The original idea of how the input and output distance function and duality results could be used to derive shadow prices is from Shephard (1970). The first study to explicitly show how the shadow prices can be computed using the (input) distance function and nonparametric linear programming methods, and demonstrating this with an application appears to be Färe, Grosskopf and Nelson (1990). In the context of shadow prices of undesirable outputs and parametric model, the approach was first applied in Färe et al. (1993). The approach is outlined also in the textbook by Färe and Primont (1995). Below,

the analytical derivation of shadow prices of outputs is presented. The presentation follows closely the exposition of Färe et al. (1993) and Färe and Primont (1995).

As shown above, the definition of an output distance function in terms of the output set $P(x)$ is

$$(8) \quad D(x,y) = \min_{y}[\theta: y / \theta \in P(x)].$$

The scalar which solves *(8)* is thus the smallest scalar by which we can divide the output vector y and still produce the implicit output vector with the given input vector x. Using the output set concept, we can analogously define the revenue function as (Färe et al. 1993, Färe and Primont 1995)

$$(9) \quad R(x,r) = \max_{y}[ry: y \in P(x)],$$

where r denotes output prices. The revenue function describes the maximum revenue that can be obtained from the given technology at output prices r, and it also completely describes the production technology. 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:

$$(10\ a) \quad R(x,r) = \max_{y}\{ry: D_O(x,y) \leq 1\}$$

$$(10\ b) \quad D_O(x,y) = \max_{r}\{ry: R(x,r) \leq 1\}.$$

The duality theorem shows that 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. In order to utilize the above duality (10 ab) in deriving the shadow prices of outputs, one has to assume that the revenue and distance functions are differentiable. We express the optimization problem in terms of the Lagrange function, i.e. the Lagrangian for *(10 a)* is

$$(11)\quad \max \Lambda(y,\theta) = ry + \theta(1 - D_O(x,y)).$$

From this we obtain the first order condition,

$$(12)\quad r - \theta \nabla_y D_O(x,y) = 0.$$

Färe et al. (1993) and Färe and Primont (1995) show that

$$(13)\quad \theta(x,r) = R(x,r).$$

In other words, the optimal value of the Lagrangian multiplier is equal to the value of the (maximal) revenue function for each (x,r). Now, define

$$(14)\quad r(x,y) = \nabla_y D_O(x,y).$$

Then, combining *(12)* - *(14)* we obtain

$$(15)\quad r = R(x,r) \cdot r(x,y) = R(x,r) \nabla_y D_O(x,y).$$

Since y is the revenue-maximizing output vector given the output price vector r and the input vector x, r can be interpreted as the (relative) output shadow price vector for y given x. Following the analysis of Shephard (1970) and Färe et al. (1993), it can be shown that the *revenue deflated relative* output shadow prices (r_m^s) for each observation can be derived as the derivative of the distance function (using dual Shephard's lemma), i.e.

$$(16)\quad r_m^s(x,y) = \partial D_O(x,y) / \partial y_m = r_m^s(x,y) / R^*(x,r^s), \quad m = 1,2,.....,M,$$

where $R^*(x,r^s)$ denotes shadow revenue. If one wishes to derive *absolute* shadow prices, instead of the above relative shadow prices, the problem is that the revenue function depends on the shadow prices r that we wish to derive. One solution to this problem is to use the following assumption (following Färe et al. 1993): *One observed*

output price equals its absolute shadow price. This assumption implies that at least one output market is efficient 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 maximum revenue is then given by

$$(17)\quad R^*(x, r^s) = r_1^o / (\partial D_o(x,y) / \partial y_1).$$

The absolute shadow prices for each observation of undesirable outputs ($m = 2,....,M$) can now be computed as

$$(18)\quad r_m^s = R^*(x, r^s) \bullet (\partial D_o(x,y) / \partial y_m) = r_1^o \bullet \frac{\partial D_o(x,y) / \partial y_m}{\partial D_o(x,y) / \partial y_1}.$$

In equation *(18)* the ratio of output shadow prices reflects the relative *opportunity costs* of the outputs, i.e. they are equivalent to the *marginal rate of transformation*. It may be noted that for the shadow prices of bad outputs the above expression does not require information on regulatory constraints. This is important because we often do not have data on regulations, and even when such data exists, the plants rarely operate exactly at the level of the constraint. 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). Further, the shadow prices do not require that the plants operate on the production frontier. Since the output distance function is homogeneous of degree +1 in outputs, the derivatives, which give the shadow prices, are homogeneous of degree zero with respect to proportional scaling of outputs. Since the output distance function is such a proportional scaling of outputs, the shadow prices are independent of whether the observations are on the frontier (see Grosskopf and Hayes 1993).

The result of equation *(18)* is also illustrated in *Figure 3*, in which is shown the output set *P(x)* that is consistent with weak disposability of bad output (pollution). Consider three possible points: A, B and C. Each of them is below the frontier (i.e. production is technically inefficient), and one cannot compute the shadow prices at these

points, since there is no tangent hyperplane 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 *(18)* 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 the bad is negative, at point B* it is zero, and at point C* positive. The plant may be operating at point A*, for example, because of environmental regulations. At point B*, the bad does 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*).

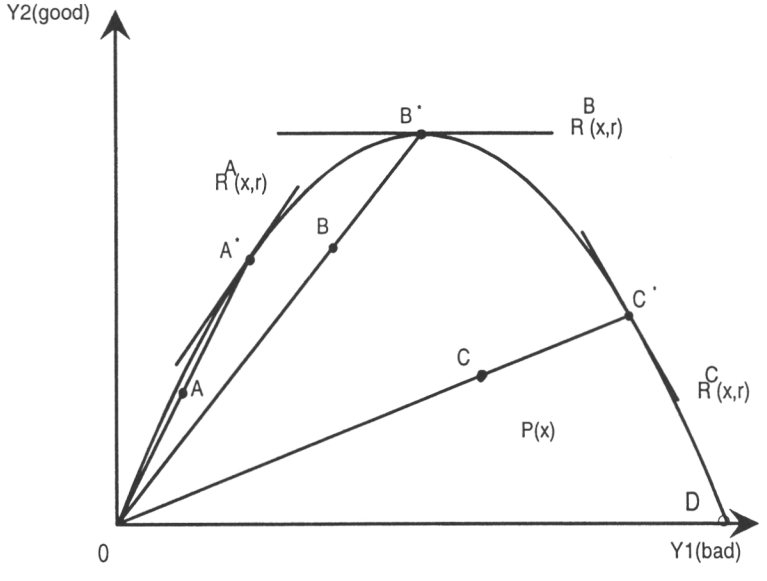


Figure 3. Technology P(x), Revenue R(x,r) and Shadow Prices of Bads

2.2.4 Empirical Applications of Distance Functions

In the empirical applications of distance functions, the deterministic linear programming method has so far been the most commonly used. The linear programming studies can be classified into two categories: nonparametric and parametric. As noted earlier, econometric or stochastic applications of the distance functions have been rare. Indeed, Lovell et al. (1990, footnote 6) state that *"Although empirical computation of distance functions using econometric techniques is certainly in its infancy, mathematical programming techniques have been used to calculate distance functions for many years now"* . Also, the recent textbooks by Färe and Grosskopf (1994) and Färe and Primont (1995), which include a number of examples of how to compute distance functions, do not provide any discussion of stochastic (econometric) estimation of distance functions. Below, we present a short summary of the empirical applications of distance functions, with an emphasis on applications related to the topic of the present study.

2.2.5 Nonparametric Linear Programming Studies

The word nonparametric simply means that there is no parametric functional form (except piecewice linearity) imposed for the model. The most common method of computing the nonparametric models is linear programming (activity analysis). The first applications of distance functions were nonparametric linear programming models. Although, there were some formulations of efficiency analysis in terms of distance functions and nonparametric linear programming in the 1970s, the empirical applications in extensive form did not start until the mid-1980s. The early studies used distance functions exclusively to analyze production efficiency (Färe, Grosskopf, and Lovell 1985).

It may be noted that simultaneously (or somewhat later) with the growth of the literature on efficiency measurement using distance functions, the operations research and management literature came up with the *Data Envelopment Analysis (DEA).* DEA provides a similar (activity analysis) approach to efficiency measurement as to the nonparametric distance function approach. The major difference is that distance functions are firmly based on neoclassical production theory, whereas no such theory lies behind the DEA. Thus, in DEA one does not define the production technology. Moreover, there is no parametric functional representation of DEA.

The study by Färe, Grosskopf and Lovell (1985) was the first popular and extensive exposition of how distance functions could be used to measure production efficiency. The book by Färe, Grosskopf, and Lovell (1994) gives a number of numerical examples of different ways to apply nonparametric computation of distance functions. The examples are mostly related to the measurement of efficiency and productivity change.

The studies by Färe et al. (1989), Yaisawrang and Klein (1994), Hakuni (1994), and Brännlund et al. (1995) are typical examples of the nonparametric approach using the weak disposability assumption to compute the impact of environmental regulations on production efficiency. To put it simply, these studies are based on the idea that one computes the linear programming problem and thereby the production efficiency under two different sets of restrictions: first by ignoring the undesirable outputs (or with strong disposability of undesirable outputs), and secondly, by incorporating bads explicitly in the optimization problem (and imposing weak disposability of bads). The difference between the two sets of efficiency scores is then interpreted as the measure of pollution control or regulation.

One of the disadvantages of the nonparametric approach is the limited amount of information it provides. In particular, for the present purpose it has the drawback that it does not directly provide information about the shadow prices of outputs (or inputs).[6] The nonparametric frontier is piecewice linear and, thus, non-differentiable at the corners (where the efficient observations are). Perhaps an even more important drawback of the nonparametric approach is the fact that it is deterministic. Because it does not allow for stochasticity and random errors, the efficiency scores may be confounded e.g. with measurement errors, and inference about the consistency of the "estimates" is difficult. However, it may be noted that recently there have been attempts to establish statistical properties for frontier models based on nonparametric linear programming (see e.g. the articles by Banker 1996, Grosskopf 1996, Seiford 1996, and Simar 1996). The current state-of-art in this literature appears to be that statistical inference can indeed be conducted in DEA-type of models, but that the statistical properties are still hard to derive and only asymptotic. In his review of the DEA literature, Seiford (1996) concludes that *"development of stochastic DEA which can incorporate measurement error and other sources of noise that inevitably contaminate the data used in an analysis is far from complete"* (p. 107).

2.2.6 Parametric Linear Programming Studies

In very recent years, the distance function approach has also been used for the purpose of computing the shadow or virtual prices for inputs and outputs. The bulk of these applications have been based on the translog linear programming (LP) model computed using the Aigner and Chu (1968) method (e.g. Färe et al. 1993, Althin 1994, Grosskopf, Margaritis and Valdmanis 1995, Coggins and Swinton 1996).

When the LP method is used for logarithmic functional forms, the "fitting criterion" is usually to minimize the sum of *absolute* or *squared deviations* of the observations below the frontier, where the deviations are defined as the logarithms of the inverse of the distances, i.e. $min\left[(ln1) - ln(D_O)\right]$. It is customary in the literature to interpret this deviation or "error" term as the reciprocal of the Farrell output-based technical efficiency index. Thus, the method does not allow for random errors and the deviations from the frontier are interpreted purely as reflecting production inefficiency (Greene 1993 a,b). This approach is presented in more detail in Essay I.

The impetus for these studies originated mainly from two class of studies, namely, studies analyzing public sector performance (Grosskopf and Hayes 1993, Althin 1994, Grosskopf et al. 1995, 1996), and studies analyzing the impact of undesirable outputs (pollution) on firms' production (Färe et al. 1993, Hetemäki 1994 a,b, 1995 a,b, Brännlund 1996, Coggins and Swinton 1996). The great virtue of the distance functions in these applications is that the duality results established for distance functions (discussed above) allow one to derive shadow prices for inputs or outputs that are employed or sold in non-competitive markets, or for outputs with no markets and thus no prices (e.g. pollution).

As was noted earlier, the study by Färe et al. (1993) was the first one to apply the distance function to derive shadow prices of undesirable outputs. Färe et al. Use a translog output distance function and linear programming to examine the impact of pollution control on pulp and paper plants. The purpose of their study is to construct shadow (virtual) prices for water pollution parameters and to examine the cost effectiveness of the regulatory system and its impact on firm efficiency. They assume *weak disposability* of undesirable (bad) outputs. The output distance function, which is dual to the revenue function, combined with the weak disposability assumption, enables the computation of the shadow (virtual) prices for pollution, without knowledge of the actual regulations, abatement costs, or prices. Furthermore, it allows one to identify the

shadow prices of pollution at the level of individual plant. Since Färe et al. (1993) interpret these shadow prices to reflect the impact of environmental regulations faced by the plant, they can be used to assess the effectiveness of existing regulation. Moreover, perhaps due to this interpretation, they also restrict, a prior, the shadow prices of bads to be non-positive. The data is based on a cross section of 30 paper and pulp mills operating in Michigan and Wisconsin in 1976.

One of the central results of Färe et al. indicates that the average shadow price of BOD is - $1,043. Thus, a reduction of biological oxygen demand (BOD) effluents by one ton diverts enough resources to have produced over two tons of paper (the price of paper being $480 on average). The mean values of the shadow prices range from zero (the absolute minimum of the model) for suspended solids to an average of -$25,270 for particulates.

More recently, exactly the same approach as in Färe et al. (1993) has been applied by Coggins and Swinton (1996). They compute the shadow prices of SO2 abatement of Wisconsin coal-burning utility plants. Again, the shadow prices are restricted to be non-positive

The advantages of the deterministic linear programming approach are that it 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 for 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 important weakness for the present purpose is that the method provides no information concerning the precision of the shadow price "estimates". Finally, the efficiency of the parametric linear programming estimators is an open question, since expressions for their asymptotic covariance matrices have never been derived (Greene 1993 a,b).[7]

2.2.7 Econometric Studies

It may be a rather strong assumption to claim that none of the possible sources of errors are present in the computation of distance functions, as assumed in the linear programming approaches. Of course, in some cases it may turn out be that the errors

are of negligible importance for the final results, but even in this case, this is usually not known a priori. Consequently, it would be important to be able to estimate the stochastic (econometric) distance function. As it turns out, however, the estimation of distance functions are not as straightforward as the estimation of conventional production, cost, or profit functions. Indeed, this may be the reason that so few econometric applications of distance functions have thus far been published.

Consider the stochastic distance function formulated in equation *(19)*,

$$(19) \quad D_O = f(X,Y;\alpha,\beta) + \varepsilon \ ,$$

where D_0 is the distance measure, *f(.)* is the production technology, X is a vector of inputs, Y is a vector of outputs, α, β are vectors of parameters to be estimated, and ε is the additive error term. The error term may be generated for various reasons. Typically it may include errors introduced by measurement, data collection, functional form specification, computational procedures, or factors known to the production units but not to the econometrician. A detailed analysis of the different factors that can generate random errors in production models and a description of the models that specify the stochastic process as an integral part of the underlying model of the firm are presented e.g. in Fuss, McFadden, and Mundlak (1978) and more recently in Brown and Walker (1995) and Griliches, Z. and J. Mairesse (1995).

The basic problem with distance functions, as concerns econometric estimation, is that one does not 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. $D_O = ln(1) = 0$).

In order to avoid the above problems, Lovell et al. (1990), Grosskopf et al. (1996) and Coelli and Perelman (1996) utilize the property that the output distance function is homogeneous of degree +1 in outputs (see also Grosskopf and Hayes 1993). Thus, for each observation to be used in estimating the distance function, a value that is unique to that observation can be used to multiply all output values on the right hand-side and the value of the distance function on the left-hand side. Thus, for an output distance function the following relationship (ignoring the error term) holds:

(20) $D_o(X,\lambda Y) = \lambda D_o(X,Y)$, for any $\lambda > 0$.

Typically, in the literature one of the outputs is chosen arbitrarily as a scaling variable (Lovell et al. 1990, Grosskopf et al. 1996, Coelli and Perelman 1996). For example, if we chose the M-th output, and set $\lambda = 1/Y_M$, equation *(20)* may be written as

(21) $D_o(X,Y/Y_M) = D_o(X,Y)/Y_M$.

Now, assume that we impose some logarithmic functional form on the output distance function, in accordance with most of the empirical literature. Then, equation (21) becomes

(22) $ln(D_o/Y_M) = f(X,Y/Y_M,\alpha,\beta)$,

where *f* denotes some logarithmic functional form, such as translog and α,β the parameters. Alternatively, equation (22) may be expressed as,

(23) $ln(D_o) - ln(Y_M) = f(X,Y/Y_M,\alpha,\beta)$

or

(24) $-ln(Y_M) = f(X,Y/Y_M,\alpha,\beta) - ln(D_o)$.

Given the data, the parameters in equation *(24)* can be estimated in various ways, depending on the estimation criteria chosen. Basically, the objective for the estimation method is to generate parameter estimates that fit the data as closely as possible while maintaining the requirement that $0 < D_O \leq 1$, which in the logarithmic case implies $-\infty < ln(D_O) \leq 0$.

In the literature, basically four different methods has been used to estimate the parametric output distance function: 1) the *linear programming (LP)* method of Aigner and Chu (1968), discussed above 2) econometric estimation using *Corrected Ordinary Least Squares (COLS)* (e.g., Lovell et al. 1990, Grosskopf et al. 1996, Coelli and Perelman 1996), 3) the Panel Data Methods without correction for the intercept (PDM) (Hetemäki 1994 a,b, 1995 a,b), and 4) the stochastic frontier model using *Maximum*

Likelihood (ML) estimation (Grosskopf and Hayes 1993, Coelli and Perelman 1996). Clearly, the most popular method has been LP (1), and the least often applied the ML (4).

The COLS approach is related to the problem that the value of the output distance function, in theory, should never exceed that for plants operating on their frontier, i.e. one. However, in the estimation of a stochastic distance function, an error term with mean zero but positive variance is assumed. For some plants the forecasted value of the output distance function can therefore exceed the theoretically plausible value. To account for this problem, the COLS method may be used (Lovell et al. 1990, Grosskopf et al. 1996, Coelli and Perelman 1996). This amounts to calculating first the most negative residual from the estimated output distance function and then adding that residual to the intercept term so that the corrected estimates of the output distance function never exceed the theoretically plausible value for any plant. In other words, this ensures that all observations are enveloped from above.

The essence of the PDM is the idea that the "fixed effects" (FE) or "random effects" (RE) from the panel data literature are interpreted as the one-sided inefficiency term of the frontier literature. The fixed effects approach is based on Ordinary Least Squares (OLS) estimation and the random effects approach on Generalized Least Squares (GLS) estimation. In the FE approach one does not assume a functional form for the fixed (inefficiency) effects, and if there are no time-invariant inputs it is not necessary to assume that these effects are distributed independently of the right hand side variables (inputs and/or outputs). The RE approach makes no distributional assumption on the inefficiency term but rather assumes that the inefficiency effects are uncorrelated with the right- hand side variables (for more details, see Appendix III). The fixed or random effects may be interpreted as inefficiency factors. Using the COLS-type of correction for the fixed or random factors one can ensure that the frontier is enveloped from above. However, if one is not interested on the production efficiency aspect but only on the shadow prices, this type of correction is not necessary since the shadow prices are not effected by the correction (i.e. to the change of the value of the intercept).

Coelli and Perelman (1996) have stressed that there is an important difference between estimating a production function formulation with $-lnY_M$ as the dependent variable by OLS and estimating an output distance function that has been scaled by lnY_M, like equation *(24)*, by OLS. The conventional OLS method on estimating a production function minimizes the sum of the squared deviations between observed and

predicted values of lnY_M. However, for the output distance function (like equation *(24)*), the fitting of a function is computed by minimizing the sum of the squared (log) *radial deviations* between observed and predicted values of the natural logarithm of the *norm* of the output (Y) vectors. In order to emphasize the difference between these two approaches, Coelli and Perelman (1996) label the latter method as *Euclidean Least Squares (ELS)* method.

Although both the COLS and FE approachs provide standard errors of the estimates and include random errors, thus allowing statistical inference, the computation of inefficiency in these approaches reverts the analysis essentially back to the deterministic frontier model. One way to avoid this drawback is to use the stochastic frontier ML method. The stochastic frontier ML method was first used by Aigner et al. (1977) in a production function context. The approach is based on the composed error term idea, in which a symmetric error term accounts for noise and an asymmetric error term accounts for production inefficiency. For the inefficiency component of the error term, one assumes a functional form and simultaneously estimates all the technology parameters and the parameter(s) of the distribution of the inefficiency term. Adding a symmetric error term, v, to equation (24), and denoting the distance to frontier term, $-ln(D_O)$, by μ, we obtain the stochastic frontier output distance function

$$(25) \quad -ln(Y_M) = f(X, Y/Y_M, \alpha, \beta) + v + \mu .$$

Typically, in the literature it has been assumed that v is distributed $iid \sim N(0, \sigma_v^2)$ and independently from μ, while μ is assumed to be either half-normal, truncated normal, exponential, or gamma distributed (see Greene 1993 a,b). It appears that the most popular choice for application has been the half-normal distribution and ML estimation (Coelli 1995).

Let us assume that we have panel data with observations of a cross section of plants (p) over time (t), i.e. p = 1.....P plants, and t = 1....T time periods. The predicted value of the output distance function for the p-th plant and t-th time period is then $D_O = exp(-\mu_{pt})$. The practical problem is that μ_{pt} is not directly observable, since it only appears as part of the composed error term, $\varepsilon_{pt} = v_{pt} + \mu_{pt}$. However, as originally suggested by Jondrow et al. (1982) and later by Battese and Coelli (1988) in a panel data context, the predictions of μ_{pt} can be obtained using conditional expectation

formulae. For the logarithmic function with a half-normal inefficiency distribution, the predicted value of the output distance function for the *p*-th plant in the *t*-th time period is (c.f., Coelli and Perelman 1996)

$$(26)\quad D_{opt} = E\left[exp(-\mu_{pt})\middle|\varepsilon_{pt}\right] \text{ or}$$

$$(27)\quad D_{opt} = \frac{1-\Phi\left(\sigma_A - \gamma\varepsilon_{pt}/\sigma_A\right)}{1-\Phi\left(\gamma\varepsilon_{pt}/\sigma_A\right)} \bullet exp\left(\gamma\varepsilon_{pt} + \sigma_A^2/2\right),$$

where $\Phi(\cdot)$ represents the cumulated distribution function of a standard normal variable, $\sigma_A = \sqrt{\gamma(1-\gamma)(\sigma_\mu^2 + \sigma_v^2)}$, and $\gamma = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_v^2)$. The predictor of D_{opt} is obtained by replacing the unknown parameters of the function *(27)* with their ML estimates.[8]

It should be noted that in the literature there appears to be some ambiguity about the impact of the homogeneity transformation of the above distance function on the estimation method and robustness of the results. After the transformation, the multiplicative (output) variable appears on both the left -and right-hand sides of the equations, which may result in endogeneity on the right-hand side. Atkinson, Färe and Primont (1996) have argued that this transformation (and the fact that some of the right hand-side variables are random) necessarily leads to simultaneity bias and heteroskedasticity problems, thus suggesting that one should use the instrumental variables estimation method and correction for heteroskedasticity. By contrast Coelli and Perelman (1996), for example, have argued that this type of transformation does not necessarily make the right-hand side variable endogenous. The argument is that if one transforms the left-hand side and right-hand side output (or input variables in the case of an input distance function) variables of an output distance function with a norm, *II y II* (*y* being one of the outputs), then the left hand side variable , *1/ II y II*, can be thought of as a choice (endogenous) variable, but the right hand side variable, *y/ II y II*, i.e. the **output mix**, can be considered fixed (exogenous). Or in the words of Coelli and Perelman (1996), *"only ratios of the outputs appear as regressors and these ratios may be assumed to be exogenous, since the distance function is defined for radial (proportional) expansion of all outputs, given the input levels, and hence by definition the output ratios are held constant for each firm"* (p. 10).

Atkinson, Färe, and Primont (1996) do not discuss the issue that the potential problems caused by the homogeneity transformation, as well as the testing for such problems, may depend on the type of distance function model and estimation method used. For example, the discussion in Davidson and MacKinnon (1993, chp. 14) indicates that generally the potential problems followed by transforming (e.g. using the Box-Cox transformation) the dependent variable and the regressors are less severe if maximum likelihood estimation is used rather than ordinary least squares. Moreover, in the context of the conventional stochastic frontier model of the type of Aigner et al. (1977), the instrumental variable or generalized methods of moments (GMM) estimator may be of rather small importance if one is interested only in the inefficiency term. Greene (1993 a, p. 80) notes that the GMM estimators for these models *"provide a high level of generality, but at the very high cost that the method produces no definable estimates of* μ_{pt}. .... *As such, the appeal of a GMM estimator of the slope parameters is intellectual at best."*

Moreover, it is not clear how one should test for endogeneity bias and heteroskedasticity in the Aigner et al. (1977) type of model, where the error term is composed of two different parts (symmetric and asymmetric). The issue appears to become even more complicated in the case of a panel data stochastic frontier model which is estimated simultaneously with a model for determinants of inefficiency, using the maximum likelihood method (as in Esseay III). Indeed, regarding the latter type of model (i.e. in the Battese and Coelli 1995 type of model), one can argue that the potential heteroskedasticity problem is partly handled implicitly. The stochastic frontier model includes the model for technical inefficiency, which is in fact a model for the second part of the composed error term. Since the variance of the inefficiency term is allowed to be different for different units (e.g., pulp plants observed in different years in Essay III) and is a function of a number of explanatory variables (constant, regulation, capacity utilization rate in Essay III), the model does implicitly account for (explain) heteroskedasticity in the error term. However, it may not account for all possible factors - and it does not account for the possibility of heteroskedasticity in the white noise error term.

The approach taken in the present study regarding the potential problems of endogeneity and heteroskedasticty is a pragmatic one. When possible, we test whether simultaneity bias and heteroskedasticity are problems in the estimations. If this is the case, instrumental variables estimation with correction for heteroskedasticty should be used. Moreover, it appears natural to treat the transformation variable, i.e. $\lambda = 1/Y_M$, as a stochastic (random) variable, thus transforming the dependent variable into a

stochastic variable. Obviously, the non-stochastic right-hand side variables which are transformed also become stochastic variables. However, as is well known, the results derived for *non-stochastic* right hand side variables (at least as regards finite sample results, such as unbiasedness and the Gauss-Markov theorem and the conventional test statistics, such as t, F, and chi-squared) hold also when the right-hand-side variables are *stochastic* (e.g. Judge et al. 1985, sec. 5.4). What is critical is whether this transformation effects the independence of the regressors from the disturbance term. If the regressors are no longer independent of the error term, instrumental variable estimation should be used. Thus, one should test for independence using e.g. the *Hausman test* (see e.g. Davidson and MacKinnon 1993).

Although there have so far been only a few econometric applications of the distance function, the variety of applications and subject areas studied is already impressive. This is perhaps an indication of the flexibility of the approach and a sign that one would expect to see a rapidly increasing number of econometric distance function studies in the future. The econometric applications so far include the following: Grosskopf and Hayes (1993) examine whether bureaucrats behave according to an economic optimizing principle (as cost minimizers, in particular) using a generalized Leontief input distance function to study the input choices of 154 Illinois municipal governments; Lovell et al. (1990) use input and output distance functions to study standard of living and inequality of the distribution of income in Australia; Grosskopf et al. (1996) use a cost indirect input distance function to model school district production technology and performance (efficiency) in Texas; Coelli and Perelman (1996) use input and output distance functions to examine the performance of European railway companies; and finally Atkinson et al. (1996) use an input distance function to measure the technical and allocative efficiency of US airlines. It appears that the only econometric applications of distance functions which analyze the impact of pollution control and environmental regulations on producer behavior are Hetemäki (1994a,b, 1995a,b) and the unpublished study by Brännlund (1996). The three essays of the present study also fall into this category of studies.

2.2.8 Concluding Remarks

It was in the early 1980s that the empirical literature on the effects of pollution control and environmental regulations on producer behavior really got started. The bulk of the literature is based on production, cost, or profit functions, modified to include abatement capital or environmental regulation intensity variables as additional factors besides the conventional inputs. The vast majority of the studies are based on the assumption that pollution control and regulations have negative impacts on productivity and production costs. Therefore, it may come as no surprise that the empirical results have more or less confirmed this assumption.

In very recent years, there has developed an interest in using the distance function approach to represent production technology. This is perhaps not surprising but rather reflects the increased demand for such functions. Indeed, which type of empirical functions have been popular in modeling producer behavior has to some extent reflected the requirements of topical issues in the society. For example, the cost and profit functions in empirical applications of producer behavior became particularly popular after the mid-1970s. Although this tendency partly reflected new theoretical results related to the duality of production, cost and profit functions, there were at least two other important factors behind the new interest in cost and profit functions.

First, it was a matter of data availability. As Shephard noted already in (1953) *"Statistical studies of cost functions are generally more accessible than corresponding empirical investigations of production functions, because economic data are most frequently in price and monetary terms" (p.* 28). Secondly, the oil crises shocks of 1973 and 1979 brought into picture the importance of energy prices and price data in general. Thus, the need to study the impact of dramatic changes in raw material prices created a demand for cost and profit functions. The common issue these studies analyzed was the substitution possibilities between different inputs. The classical example of this strand of the literature is the study by Berndt and Wood (1975).

Societal changes in the 1980s and 1990s have placed new demands on production models which emphasize the need to use more general models and - somewhat curiously - the need to go back to quantity data. First, there has been a growing concern about environmental issues and on the impact of regulations on firm and industry performance. This interest has generated a new challenge to econometric production models. For example, as discussed earlier, price data (such as abatement expenditures)

are often notoriously bad or simply not available. On the other hand, bad outputs (pollution) are not generally sold on markets (except in case of tradable permits); thus these outputs do not have market prices (data).

On the other hand, the problems concerning growing public sector deficits in the western industrialized countries has forced governments in many countries to rationalize the public sector. This in turn has created an interest in studying the efficiency of public sector organizations, such as hospitals, schools, police force, railways, etc. Typically, in these sectors the conventional behavioral hypothesis of cost minimization or profit maximization does not hold. These sectors are often operating under budget constraints and in imperfect markets (e.g. monopolistic) and their outputs and/or inputs may not have market prices. Moreover, these sectors probably fail to operate on production frontiers (i.e. technically efficiently) more often than firms in the competitive markets.

As has been noted earlier, the above features put special requirements on production models, which the conventional production, cost and profit functions do not fully satisfy. In contrast, distance functions can be, and have been, a solution to these problems. In light of the above background, it appears only natural that distance functions have recently become popular in the literature. However, it seems that the applications of distance functions is still in the expansive stage of the "production cycle", and one would expect to see a rapidly increasing number of applications in the near future. Moreover, the fact that distance functions are more general representations of production technology than the conventional production, cost or profit functions suggests that the interest in distance functions is going to be a lasting one.

3. Summary of the Essays

The essays of the present study contribute to the empirical methodology of computing distance functions and to the substantive issues related to the impact of pollution control and environmental regulations on the pulp industry. The methodological contributions are of three kinds. First, in *Essay I*, the deterministic parametric linear programming (LP) approach and the econometric approach to estimating a distance function are compared for the first time. All the previous studies have used one or the other of these methods (most using LP), but no study has tried to analyze how sensitive the results are to the particular approach chosen.[9] Moreover, the parametric linear programming studies have to a large extent taken the results at their face value. That is, sensitivity analysis using various model specifications (e.g. functional forms) or diagnostic checking have not been applied. Indeed, this is perhaps not surprising, since the method itself does not provide statistics, such as standard errors, which could be used to analyze the consistency of the estimates. However, sensitivity analysis and "methodological cross-checking" using alternative approaches to a common set of data is particularly useful when important policy issues are to be guided by the results. This is the case, for example, when the results are used to help to design environmental policy. Finally, it should be stressed that in the present study particular emphasize has been put on the construction of good quality data set. Indeed, the type of *plant level panel data* (with relatively long time dimension) used in the present study is still rare in the environmental economics literature.

The results of the *Essay I* show that the computed shadow prices are sensitive to the methodology used. Moreover, the results raise the important question of the robustness of the parametric linear programming approach. For the particular case of the Finnish pulp industry, the results indicate that the stochastic approach is more stable.

Moreover, in contrast to all the previous studies, both the parametric and stochastic models used here relax the assumption that shadow prices of undesirable outputs must be non-positive. In the literature, the question of why such a restriction should be imposed a priori has been left open. Since the axioms on which the distance function is based do not require such a restriction, there is no need to set it and to rule out the possibility of positive shadow prices. The shadow price ratio reflects the marginal rate of transformation (opportunity cost) between the good and bad outputs, and, as will be

argued in this study, there are a number of reasons why one might expect this ratio at times to be positive.

In *Essay II*, a new (two-stage) method of estimating a stochastic distance function is proposed and applied to compute shadow prices for undesirable outputs (pollution). The basic problem with distance functions, as concerns econometric estimation, is that one does not usually observe (have data on) the dependent variable, i.e. the distance from each observation to the production frontier. Further, if one sets the distance function equal to its efficient (frontier) value (= 1), the left-hand side of the distance function is invariant, an intercept cannot be estimated, and ordinary least squares (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. The most common way of overcoming this problem is to set the left-hand side equal to one and utilize the property that the output (input) distance function is homogeneous of degree +1 in outputs (inputs) and to transform the equation to an estimable form. In *Essay II*, a different method is used. First, the distance from each observation to the production frontier is computed using a nonparametric linear programming model, and in the second stage, these values are used as the dependent variable in the stochastic model. Thus, the dependent variable is based on actual distance scores, and the method reduces to the conventional one only if the distance score for each observation is equal to one.[10] The essay also provides information of the sensitivity of the results to various functional form specifications and estimating approaches (OLS and a Tobit maximum likelihood method).

Essay III is concerned with estimating the impact of environmental regulation on the production efficiency the Finnish pulp plants. The essay contributes to the literature by extending the existing *stochastic frontier literature* by incorporating undesirable outputs and environmental regulations in stochastic frontier models. Moreover, it provides an empirical framework to test the *Porter hypothesis*. According to the Porter hypothesis, properly designed environmental regulations can actually stimulate production efficiency gains that more than offset the costs of complying with the regulation (e.g. Porter and van der Linde 1995). The correctness of this hypothesis is still largely an open question, due to the lack of empirical analysis (see Jaffe and Palmer 1996). It appears that no study has been done to directly test the hypothesis. We now proceed to a more detailed discussion of the above two contributions of *Essay III*.

The *stochastic frontier approach* to production efficiency measurement was pioneered simultaneously (on three different continents) by Aigner, Lovell, and Schmidt (1977), Battese and Corra (1977), and Meeusen and van den Broeck (1977). The literature was motivated by the idea that deviations from the production frontier might not be entirely under the control of the producer, as the deterministic linear programming frontier literature implicitly assumes. Instead, it was assumed that any particular producer faces its own production frontier and that frontier is randomly placed by the whole collection of stochastic elements which might enter the model outside the control of the producer. As the recent surveys by Greene (1993 a,b) reveal, the stochastic frontier approach has rapidly gained popularity and presently there exists an extensive body of empirical applications of stochastic frontier functions. However, it appears that no study has explicitly included pollution or environmental regulations in the estimation of production efficiency. The studies have been concerned with a single output (good output), not have they included pollution as an input. This may be a significant shortcoming, since a large part of the stochastic frontier literature is concerned with production processes generating substantial amounts of waste and which are regulated (e.g. power generation, pulp and paper production, base metals, steel production, etc.). Thus, the previous studies have implicitly assumed that undesirable outputs and/or environmental regulations do not have any impact on efficiency measurement. Yet, the results from nonparametric linear programming studies which have explicitly incorporated undesirable outputs in the production technology model indicate that the neglect of these factors can greatly bias the production efficiency measures (e.g. Färe et al. 1989, Yaisawarng and Klein 1994, Hakuni 1994). *Essay III* shows how undesirable (or multiple) outputs and environmental regulation can be explicitly included in stochastic frontier efficiency measurement. The method used in the present study extends the stochastic frontier model developed by Battese and Coelli (1995) in a production function setting to the distance function framework. By doing this, it also provides a method for testing the Porter hypothesis.

The substantive contributions of the essays are of the following type. First, the study is a first attempt to analyze separately the unregulated and regulated bad outputs. Indeed, the previous studies on pulp and paper industry water pollution have not included the (unregulated) total waste water flow (FLOW) as an undesirable output. However, arguably FLOW is the most important bad output variable that the plants are trying to control. Moreover, the FLOW variable is controlled mainly by pollution

prevention methods (closing the water system), whereas BOD and SS, the most common bad outputs included in the empirical studies, are mainly controlled by end-of-pipe measures (treatment lagoons) (see Appendix I). Thus, the study provides new insight on whether the impact of controlling the unregulated FLOW differs from the control of regulated bad outputs (BOD, SS). Secondly, the essays contribute to the subject in the national context by providing the first attempt to estimate econometrically the impact of pollution control or/and environmental regulation in Finland on firm or industry production performance. This may be considered to be somewhat surprising against the fact that the pulp and paper industry and the metal industry are the two most important industrial sectors in Finland and that both of these industries have been subject to environmental regulations. Consequently, one would expect that there would be a great deal of interest, e.g. amongst the regulator and the industries themselves, as to whether the control measures cause costs, and if so, how large.

The results of the essays fall basically into two categories: those related to *shadow prices* of undesirable (or bad) outputs and, those related to the impact of regulation on production *efficiency* (i.e. to the Porter hypothesis). Let us consider the shadow price issue first.

All of the three essays estimate the shadow prices of the bad outputs. Since the bad outputs do not have markets, or therefore prices, one has to resort to shadow prices. These shadow prices reflect the costs of pollution control measures to the plant. If the plant can reduce effluents only by shifting resources from the production of good output to the control process, this will result in foregone good output (revenue), for a given amount of inputs. In this case the shadow price of the bad output will be negative. However, if the effluents are in fact to any extent a result of an inefficient production process, it is possible for the plant to simultaneously reduce effluents and increase good output, with a given amount of inputs. In this case, the shadow price of the bad output is positive. Finally, if the reduction of effluents does not have any impact on the production of the good output, the shadow price is zero. Alternatively, the shadow price can be interpreted as the foregone revenue (negative) or revenue gain (positive) due to changes in the effluent discharges.

The results of all three essays indicate that the shadow price of waste water flow (FLOW) which has not been regulated is positive. Therefore, the abatement of waste water effluents does not necessarily impose additional costs to the plants but, to the contrary, may even increase the productivity of the plants. This result is interesting not

least due to the fact that the vast majority of economic analyses of the impact of pollution control on the firm's production performance are based on the assumption that these measures have a negative impact on the production of good outputs. Thus, environmental goals have been viewed to be in conflict with business goals. The positive result obtained in the present study is probably due to the fact that the recirculation of waste water and closure of the water system is associated with a decrease in material waste and therefore with a reduction of inefficiency. Indeed, the long term objective of the pulp plants has been gradually to close the water system totally. It is important to note that this development has evolved, to an important degree, independently of environmental regulations.

On the other hand, the results for the shadow price for the *regulated* bad output, namely, biological oxygen demand (BOD) are more ambiguous. Depending on the model specification, the shadow prices are either positive or negative. However, in most of the cases it was negative. A negative shadow price for BOD would indicate that the opportunity cost of reducing BOD has been foregone revenues. The likely reason for this is that the abatement of BOD has been achieved to a great extent through building external treatment plants (aerated lagoons), besides recirculating the waste water. These treatment plants impose additional costs on pulp production without having any positive impact on the amount of pulp produced.

The above results are to some extent contrary to studies that have analyzed the impact of pollution control on US pulp and paper mills (e.g. Färe et al. 1993). This may be due either to a difference in Finnish and US pulp mill production technology, or to model restrictions, or to the data sets. As was indicated above, the previous studies have a prior restricted the shadow prices to be non-positive. The results from the present study indicate that it is important not to restrict the shadow prices of undesirable outputs to be non-positive.

The second substantive result concerns the impact of regulation on production efficiency. The results of *Essay III* show that an increase in regulation intensity (standard) has led to a decrease in the production efficiency of the pulp plants. The results therefore do not appear to give support to one of the central premises of Porter's hypothesis, but rather agree with the conventional economic textbook model, whereby an increase in the stringency of environmental regulations unambiguously makes the polluting firm worse off. Thus, the environmental standard on BOD (and suspended solids), as such, has a negative impact on production efficiency, although the abatement of FLOW tends to

increase production efficiency. At a more practical level, the results indicate that the water quality authorities may have failed in their design of regulations that would promote the use of innovative measures through which plants could have realized overlooked opportunities to increase efficiency. Indeed, the above results suggest that the regulator should control directly the FLOW parameter, instead of BOD and suspended solids. In other words, regulation should be targeted at pollution-prevention measures (closing the water system), rather than end-of-pipe measures (e.g. biological treatment lagoons). In relation to the Porter hypothesis, the positive shadow price of FLOW also raises a question: If the water authority had regulated FLOW instead of BOD and SS, would the impact of the regulation on production efficiency have had been positive rather than negative, as the Porter hypothesis suggests?

Finally, we would to emphasize the advantages of the data set used in this study. Typically, in economic literature matters related to the processing, quality and form of the data are not considered to be in a category of contributions. As Schmalensee (1990) has aptly noted, *"Economists, unlike historians or anthropologists, are formally trained only in the analysis of data sets, not in their construction. The economics profession does not much reward the tedious labour necessary to construct sound and interesting data sets."*(p. 163). However, in an empirical study, the data constitute the essential building block for all the work that follows. Consequently, the results are also conditional on the robustness of the data set. In the present study particular care was attached to the collection and construction of the data set.

All three essays are based on plant level panel data. The plant (or firm) level is the level which neoclassical production theory claims to comprehend. The data set consists of observations on eight sulfate pulp plants observed annually from 1972-90. At least for three reasons, the data set can be considered to be of rare quality, even internationally. First, systematic, good quality, long panel data sets rarely exist on undesirable (pollution) outputs and environmental regulation variables. Secondly, due to a lack of reliable micro level data on these variables, studies analyzing the impact of environmental regulation have generally used industry or country level aggregate data. However, if pollution regulations are set at the plant level, it is also important to be able to measure the effects of pollution control at this level. Firm or plant level regulations are particularly common for point source waste water effluents. For example, in the pulp and paper industry in Finland and Sweden, every plant is individually regulated with respect to the substances that it discharges. Similar practices are also common in many industries in the United

States and Canada. Besides the "conventional" aggregation bias, using industry level data in such cases causes identification problems and measurement errors.

The third important feature of the data set relates to the actual data processing. 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 data collection process. 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 precision and measurement errors. In the present study, the data collection was done in close cooperation with the water authority, the pulp plants involved in the study, and the Central Statistical Office in order to minimize these errors. Not only did these parties help to correct the errors that existed in the official data, but they also contributed to the interpretation of the data, which was necessary to enable the forming of a long continuous panel.

Notes

1. It would be more accurate to talk about the impact of *effluent* control or *residuals management*, rather than the impact of *pollution* control on a firm. The word pollution should strictly be used in relation to the impact that the waste water effluent has on receiving waters. However, due to the common usage of word pollution in the present context and for simplicity, it also used here.
2. According to Pittman (1983) *"... clearly the most difficult and challenging task is likely to be the assigning of shadow prices for undesirable outputs. Even when there are explicit engineering or econometric estimates available for a particular industry or area, these are likely to be subject to a wide range of error. Where exactly appropriate estimates are not available - as is especially likely for a sample of individual plants - then, of course, an additional error is imposed"* (p. 887).
3. Brännlund and Löfgren (1996) analyze whether the water pollution regulation scheme implemented in Swedish pulp plants has been efficient. The study is motivated by the fact that the plant emissions are often observed to be below the regulation standard.
4. Uimonen (1986) experimented with different types of empirical models in an attempt to quantify the impact of environmental regulations on the pulp and paper industry. Hakuni (1994) used nonparametric linear programming to compute the efficiency of pulp plants under two regimes. First, he computed a model in which no undesirable outputs were included in the analysis. These "baseline" results were compared to results obtained from a model, in which water pollution "outputs" were explicitly included in the analysis. The difference in the production efficiency scores from the two models were interpreted to describe the impact of undesirable outputs on efficiency.
5. The first six axioms for the output set in Färe (1988 p. 6) are the following:

   $P.1 \quad (a)\ 0 \in P(x),\ \forall\, x \in R_+^n.$
   $(b)\ y \notin P(0),\ y \geq 0,\ y \neq 0.$

   (a) states that inaction is possible with every input, while (b) says that there is no free lunch, i.e. inputs are required to produce outputs.

   $P.2 \quad \forall\, x \in R_+^n,\ P(x) \subset P(\lambda x),\ \lambda \geq 1.$

   $P.2.S \quad \forall x, y \in R_+^n,\ P(y) \subset P(x),\ x \geq y.$

   P.2 states that if inputs are proportionally increased, outputs do not decrease (weak disposability of inputs), while P.2.S says that outputs do not decrease if some inputs are decreased (strong disposability of inputs).

   $P.3 \quad \forall\, x \in R_+^n,\ y \in P(x)$ *only if* $\Theta y \in P(x),\ 0 \leq \Theta \leq 1.$

   $P.3.S \quad \forall\, x \in R_+^n,\ y \in P(x)$ *only if* $v \in P(x),\ 0 \leq v \leq y.$

   P.3 states that if an output vector y is obtainable with an input vector x, then all output vectors which are proportionally less than y, i.e. $\Theta y,\ 0 \leq \Theta \leq 1,$ are also obtainable with x (weak disposability of outputs). P.3.S says that if y is obtainable with x, then so are output vectors v that are smaller than y i.e. one can freely dispose of outputs (strong disposability of outputs).

P.4 $P(x)$ *is bounded* $\forall\, x \in R_+^n$.

That is, finite inputs yield only finite outputs.

P.5 $P: R_+^n \rightarrow 2^{R_+^m}$ *is a closed correspondence, i.e.*
if $[x^0 \rightarrow x^1,\ y^1 \rightarrow y^0$ *and* $y^1 \in P(x^1),\ ]$ *then* $y^0 \in P(x^0)$.

That is, the graph of the technology is a closed set. Moreover, it implies that P(x) and the input set, L(y), are closed sets $\forall x \in R_+^n$ *and* $\forall y \in R_+^m$, respectively. The closeness of these sets guarantees the existence of efficient output and input vectors. P.4 and P.5 together imply that the output set P(x) is compact $\forall x \in R_+^n$, *i.e.* P is compact valued.

P.6 *If* $y \in P(x)$, $y \geq 0$, $x \geq 0$ *then* $\forall \Theta \geq 0$, $\exists\, \lambda_\Theta \geq 0$ *such that* $\Theta y \in P(\lambda \Theta x)$.

P.6 states that in an unconstrained environment, if an output vector is obtainable, then any scalar multiplication of it is obtainable by proportional scaling of inputs (attainability).

6. Färe, Grosskopf, and Nelson (1990) provide a framework, in which nonparametric linear programming is used for deriving shadow prices for inputs using input distance function.
7. Greene (1993 b, pp. 14-15) has summarized the problems with the linear programming approach (in the computation of production efficiency) as follows: *"While the programming procedures do produce "estimates", they have the notable disadvantage that they do not produce standard errors for the estimators, so inference is precluded. Expressions for their asymptotic covariance matrices remain to be derived. For the present purposes, the main disadvantage is that absent a more detailed specification, consistency of the estimates cannot be verified nor, as such, can consistency of the inefficiency estimates* μ_i. *The programming estimators may, however, have the virtue of robustness to specification errors in the distribution of* μ_i, *though this, too, remains to be verified."*
8. Presently it appears that there are only two computer packages that have readily incorporated the above procedure in the estimation routines*: LIMDEP 7.0* and *FRONTIER 4.1.*
9. In an unpublished mimeo Brännlund (1996) has also compared parametric linear programming and econometric output distance function approaches.
10. Recently a similar type of two-stage approach has been applied in the DEA literature (see e.g. Grosskopf 1996). However, the two-stage approach used in the present study differs from this literature in that the interest is not to analyze and explain the efficiency scores. In the literature, which is surveyed by Grosskopf (1996), the second stage econometric models are always *ad hoc*, including variables which are considered not to be directly controllable by the decision making units (DMU), but important in explaining the variations in the efficiency scores. In the present study the second stage model is based on econometric output distance function, and the interest is to analyze the shadow prices of bad outputs, not the efficiency.

References

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

Aigner, D.J., C.A.K. Lovell, and P.J. Scmidt (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21-37.

Althin, R. (1994). Shadow Pricing of Labor - A Translog Distance Function Approach, Department of Economics, University of Lund, Working Paper Series 33/94.

Atkinson, S., R. Färe, and D. Primont (1996). Measuring Technical and Allocative Inefficiency via Econometric Estimation of Distance Functions, unpublished mimeo, May 20, 1996.

Banker, R. (1996). Hypothesis Tests Using Data Envelopment Analysis, *in Olesen, Petersen and Lovell (eds. 1996).*

Barbera, A.J. and McConnell, V.D. (1990). The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects, *Journal of Environmental Economics and Management* 18, 50-65.

Battese, G. and G. Corra (1977). Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia, *Australian Journal of Agricultural Economics*, 21, pp. 167-179.

Battese, G. E. and T. J. Coelli (1988). Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data, Journal of Econometrics 38; 387-399.

Battese, G.E. and T.J. Coelli. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, to be published in *Empirical Economics, vol. 20, no. 2.*

Baumol, W.J. and W. E. Oates (1988). *The theory of environmental policy* (second ed.), Cambridge University Press.

Berndt, E.R. and D. O. Wood (1975). Technology, Prices and the Derived Demand for Energy, *Review of Economics and Statistics*; 259-68.

Brown, B. W. and M. B. Walker (1995). Stochastic specification in random production models of cost-minimizing firms, *Journal of Econometrics*; 175-205.

Brännlund, R. (1996). Estimating Shadow Prices of Undesirables, unpublished mimeo, Department of Economics, University of Umeå.

Brännlund, R., R. Färe, and S. Grosskopf (1995). Environmental Regulation and Profitability: An Application to Swedish Pulp and Paper Mills, *Environmental and Resource Economics*; 23-36.

Brännlund, R. and K.-G. Löfgren (1996). Emission Standards and Stochastic Waste Load, *Land Economics* 72(2); 218-30.

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

Christainsen, G.B. and T.H. Tietenberg (1985). Distributional and Macroeconomic Aspects of Environmental Policy, in Kneese, A. and Sweeney, J.L. (eds.*) Handbook of Natural Resource and Energy Economics, vol. I*, North-Holland.

Coelli, T. (1995). Estimators and Hypothesis Test for a Stochastic Frontier Function: A Monte Carlo Analysis, *Journal of Productivity Analysis*; 247-68.

Coelli, T. and S. Perelman (1996). Efficiency measurement, multiple-output technologies and distance functions: with application to European Railways. CREPP Working Papers 96/05, Universite de Liege.

Coggins, J.S. and J. R. Swinton (1996). The Price of Pollution: A Dual Approach to Valuing SO2 Allowances, *Journal of Environmental Economics and Management* 30: 58-72..

Conrad, K. and Morrison, C.J. (1989): The Impact of Pollution Abatement Investment on Productivity Change: An Empirical Comparison of the U.S., Germany, and Canada. *Southern Economic Journal*, January: 684-698.

Cropper, M. L. and W. E. Oates (1992). Environmental Economics: A Survey, *Journal of Economics Literature*; 675-740.

Data Resources Incorporated (1979). *The Macroeconomic Impact of Federal Pollution Control Programs: 1978 Assessment*, Report submitted to the Environmental Protection Agency and the Council of Environmental Quality.

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

Debreu, G. (1951). The Coefficient of Resource Utilization, Econometrica 22, 14-22.

Denison, E. F. (1978). Effects of selected changes in the institutional and human environment upon output per unit input, Survey of Current Business, U.S. Department of Commerece, Bureau of Economic Analysis. (January 1978).

Farrell, M.J. (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A* 120: 253-90.

Frontier 4.1. See Coelli, T. (1994). A Guide to Frontier 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, Department of Econometrics, University of New England, mimeo.

Fuss, M. and D. McFadden (eds.). (1978). *Production Economics; A Dual Approach to Theory and Applications*, North-Holland.

Fuss, M., D. McFadden, and Y. Mundlak (1978). A Survey of Functional Forms in the Economic Analysis of Production, in *Fuss and McFadden (eds.1978).*

Färe, R. (1988): *Fundamentals of Production Theory*, Lecture Notes in Economics and Mathematical Systems, vol. 311, Berlin: Springer Verlag.

Färe, R. and S. Grosskopf (1983). Measuring Output Efficiency, *European Journal of Operational Research*; 173-179.

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

Färe, R., S. Grosskopf, C.A.K. Lovell (1985). *The Measurement of Efficiency of Production.* Boston: Kluwer-Nijhoff.

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

Färe, R., S. Grosskopf, and C. Pasurka. (1986). Effects on Relative Efficiency in Electric Power Generation due to Environmental Controls, *Resources and Energy*; 167-184.

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*, 90-98.

Färe, R., S. Grosskopf, and J. Nelson (1990). On Price Efficiency. *International Economic Review* 31; 709-720.

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 D. Primont (1990), A Distance Function Approacg to Multioutput Technologies, *Southern Economic Journal;* 879-891.

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

Gollop, F.M. and M.I. Roberts (1983). Environmental Regulations and Productivity Growth: The Case of Fossil-fueled Electric Generation, *Journal of Political Economy*: 654-674.

Gray, W.B. & S.J. Shadbegian (1993). Environmental Regulation and Manufacturing Productivity at the Plant Level, National Bureau of Economic Research, Working Paper No. 4321.

Gray, W.B. & S.J. Shadbegian (1995). Pollution Abatement, Cost, Regulation, and Plant Level Productivity, National Bureau of Economic Research, Working Paper No. 4994.

Greene, W. (1993a). Frontier Production Functions, New York University, Department of Economics Working Paper Series, EC-93-20. To be published in Pesaran, H. and P. Schmidt (eds.). *Handbook of Applied Econometrics, Vol. II - Microeconometrics.*

Greene, W. (1993b). The Econometric Approach to Efficiency Analysis, in Fried, H.O., C.A. Knox Lovell & S.S. Schmidt (eds.). *The Measurement of Productive Efficiency*, Oxford University Press.

Griliches, Z. and J. Mairesse (1995). Production Functions: The Search for Identification, National Bureau of Economic Research, Working Paper No. 5067.

Grosskopf, S. (1996). Statistical Inference and Nonparametric Efficiency: A Selective Survey, *in Olesen, Petersen and Lovell (eds. 1996).*

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

Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1996). Budget Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling, *Review of Economics and Statistics*, forthcoming.

Grosskpf, S., D. Margaritis, and V. Valdmanis (1995). Estimating Output Substitutability of Hospital Services: A Distance Function Approach, *European Journal of Operations Research*; 519-533.

Hakuni, M. (1994). Tehokkuus Suomen sulfaattiselluteollisuudessa vuosina 1972-90 (Efficiency in the Finnish sulfate pulp industry in 1972-90), Metsäntutkimuslaitoksen tiedonantoja (Finnish Forest Research Institute Research Reports No. 517.

Hetemäki, L. (1994a). The Impact of Pollution Control on Firm Production Technology and Efficiency: A Stochastic Distance Function Approach, in Brännlund, R., B. Kriström, and K.-G. Löfgren (eds.). Environmental Economics, Proceedings of the International Conference held at Ulvön, Sweden, June 10-13, 1993, Swedish University of Agricultural Sciences, Department of Forest Economics, Report 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.

Hetemäki, L. (1995a). The Impact of Pollution Control on Technology and Efficiency: A Stochastic Distance Function Approach, in Boero, G. and Silberston, A. (eds.). *Environmental Economics*, MacMillan Press Ltd, Great Britain.

Hetemäki, L. (1995b). Using Deterministic and Stochastic Distance Functions to Measure the Effects of Pollution Control on a Firm, Finnish Forest Research Institute, Research Papers No. 549.

Jaffe, A.B., S. R. Peterson, P.R. Portney, and R. N. Stavins (1995). Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?, *Journal of Economic Literature*; 132-163.

Jaffe, A. B. and K. Palmer (1996). Environmental Regulation and Innovation: A Panel Data Study, National Bureau of Economic Research, Working Paper 5545, April 1996.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics* 19(2/3); 233-238.

Jorgenson, D.and P.J. Wilcoxen (1990). Environmental Regulation and U.S. Economic Growth, *RAND Journal of Economics*, vol.21,no.2; 314-340.

Judge, G. G., W. E. Griffiths, R. Carter Hill, H. Lutkepohl, and T.-C. Lee (1985). *The Theory and Practice of Econometrics,* second ed., John Wiley and Sons.

Kneese, A.V. and B.T. Bower (1968). *Managing Water Quality: Economics, Technology, Institutions*. Resources for the Future, John Hopkins Press.

Kneese, A.V., R. U. Ayres, R. C. d'Arge (1970). *Economics and the Environment: A Materials Balance Approach*. Washington: Resources for the Future.

LIMDEP . See Greene, W. (1995) Limdep Version 7.0 User's Manual, Econometric Software, Inc.

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.

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

Meeusen, W. and J. van den Broeck (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18; 435-444.

Meyers, J.G. and L. Nakamura (1980). Energy and Pollution Effects on Productivity: A Putty-Clay Approach, in New Developments in Productivity Measurement and Analysis, Kenderick (ed.), University of Chicago Press.

Muller, R. A. (1978). Econometric analysis of environmental policy: estimation of a model of the Canadian pulp and paper industry, *Canadian Journal of Economics*, XI, no.2; 263-86.

Oates, W.E., K. Palmer, and P. R. Portney (1993). Environmental Regulation and International Competitiveness: Thinking About the Porter Hypothesis, Resources for the Future Discussion Ppaer No. 94-02.

Olesen, O. B., N. C. Petersen and C. A. Knox Lovell (eds.) (1996). Special Issue: Efficiency and Frontier Analysis: Proceedings of a Workshop on State-of-the-Art and Future Research in Efficiency Analysis, *Journal of Productivity Analysis*, vol. 7, No. 2/3.

Palmer, K., W. E. Oates, and P. Portney (1995). Tightening Environmental Standards: The Benefit-Cost or the No-Cost Paradigm?, *Journal of Economic Perspectives*, vol 9, No. 4; 119-132.

Pittman, R.W. (1981). Issues in Pollution Interplant Cost Differences and Economies of Scale, *Land Economics*, 1-17.

Pittman, R.W. (1983). Multilateral Productivity Comparisons with Undesirable Outputs, *Economic Journal,* 883-891.

Porter, M. E. and C. van der Linde (1995). Toward a New Conception of the Environment-Competitiveness Relationship, *Journal of Economic Perspectives*, vol 9, No. 4; 97-118.

Portney, P. R. (ed.) (1990). *Public Policies for Environmental Protection*, Resources for the Future, Washington, D.C.

Schmalensee, R. (1990). Empirical Studies of Rivalrous Behaviour, in Bonanno, G. and D. Brandolini (eds.), *Industrial Structure in the New Industrial Economics*, Clarendon Press, Oxford.

Seiford, L. M. (1996). Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995), *in Olesen, Petersen and Lovell (eds. 1996).*

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

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

Shephard, R.W. (1974). Comment on Diewert, E. "Applications of Duality Theory", in Intrilligator, M.D. and Kendrick, D.A. (eds.), *Frontiers of Quantitative Economics vol. II*, North-Holland.

Simar, L. (1996). Aspects of Statistical Analysis in DEA-Type Frontier Models, *in Olesen, Petersen and Lovell (eds. 1996).*

Sims, W. A. (1979). The Response of Firms to Pollution Charges, *Canadian Journal of Economics* 12: 57-74.
Tietenberg, T. (1992). *Environmental and Natural Resource Economics* (3rd Edition), Harper Collins.
Uimonen, S. (1986). Ympäristönsuojelun yritystaloudellisista vaikutuksista. Vesihallituksen monistesarja, Nro 192.
Uimonen, S. (1992). *Pollution Control in the Firm and the Industry: A Theoretical Study*, Annales Academie Scientiarum Fennicae, Helsinki 1992.
Yaisawarng, S. & Klein, D.J. (1994). The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry, *Review of Economics and Statistics*; 447-460.

ESSAY I

The Impact of Pollution Control on the Pulp Industry: A Comparison of Deterministic and Stochastic Approaches

Lauri Hetemäki

Abstract

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.

I am grateful to helpful comments offered on a earlier version of this paper by Robert Chambers, Mika Hakuni, Pekka Ilmakunnas, and Jari Kuuluvainen. Also, helpful comments were received at the European Association of Environmental and Resource Economics Conference, June 17- June 20, 1995, Umeå. This research was supported by the Academy of Finland and the Nordic Council of Ministers Project "Economic Instruments in Environmental Policy: The Nordic Forest Sector". A previous version of this paper appeared in the Finnish Forest Research Institute, Research Reports series under the title "Using Deterministic and Stochastic Distance Functions to Measure the Effects of Pollution Control on a Firm".

1. Introduction

This study has two distinct objectives. First, the purpose is to analyse the shadow prices of undesirable (or bad) outputs in Finnish pulp industry. These shadow prices reflect the opportunity cost of pollution abatement to pulp plants. Secondly, the analytical goal of the study is to compare the shadow prices derived from using two different techniques, namely the parametric linear programming approach and the econometric approach. This type of "methodological cross-checking" between linear programming and econometric models on a common body of data has been advocated e.g. by Charnes, Cooper, and Sueyoshi (1988) and Ferrier and Lovell (1990). Charnes et al. argue that particularly in cases where important policy issues are concerned, it will not generally be prudent to rely on only one methodology. In the case of shadow prices of bad outputs, the results are of practical importance in designing the regulatory policy. Moreover, the comparison of the two approaches is of particular interest in the present case, since none of the previous studies have performed such a comparison.

In recent years, there has been a growing interest in the use of distance functions to derive shadow prices for inputs and outputs. Studies using distance functions to compute shadow prices of either inputs or outputs in regulated industries or services include e.g. Grosskopf and Hayes (1993), Färe et al. (1993), Althin (1994), Grosskopf, Margaritis and Valdmanis (1995), and Coggins and Swinton (1996). Most of the empirical applications of distance functions have been based on deterministic nonparametric or parametric linear programming, and only a very few econometric studies exist (Grosskopf and Hayes 1993 is one exception). 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 derived (Greene 1993). 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.

This study is a first attempt to compare the relative strengths of the parametric deterministic distance function and the stochastic distance function approaches in computing the shadow prices of outputs (or inputs). Following the approach of Ferrier and Lovell (1990) and Charnes et al. (1988), and Evans and Heckman (1988), a real world data set is used, rather than controlled or Monte Carlo data. Although a Monte Carlo type of approach would allow more general inferences to be made on the performance of the two methodologies, they require the imposition of a number of restrictions on the particular case studied. Typically, in the latter type of studies, the data set is constructed artificially for the experiment and no real data is used. This is the case, e.g. in the study by Gong and Sickles (1992), who use the Monte Carlo approach to compare the performance of the non-parametric linear programming and econometric approaches to measure technical efficiency. We chose the first approach, since in the present study the particular interest is to provide new information about the impact of pollution control on the Finnish pulp industry.[1]

In summary, in the present study the parametric linear programming and econometric 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), except for some minor changes. 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. 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.

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 presents the empirical results and their implications. Concluding remarks are given in section 5.

2. Theoretical Model

A production technology transforming factors of production $x = (x_1, x_2, \ldots, x_n) \in R_+^n$ into outputs $y = (y_1, y_2 \ldots, y_m) \in R_+^m$ can be modeled 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.[2] 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.

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 $r = (r_1, \ldots, r_m)$ and it is assumed that r can be nonpositive for bad outputs. 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,

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

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

Thus, the revenue function can be derived from the output distance function by "maximizing" revenue over output quantities, and 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 (for detailed exposition of the derivation of the shadow price, see p. 25-29)

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

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 (c.f. Porter and van der Linde 1995a). 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 3.*, in p. 29.

It may be noted that unlike the previous studies (c.f. Färe et al. 1993 and Coggins and Swinton 1996), no a prior restrictions on the shadow prices are set. Both Färe et al. (1993) and Coggins and Swinton (1996) compute the shadow prices for undesirable outputs by imposing the restriction that bad outputs can have only negative or zero shadow prices. Thus, by construction, their models cannot generate shadow prices that lie between points B and D in *Figure 3,* in p. 29. Since the shadow price for bad output, i.e. equation (2.4), does not reflect the impact of environmental regulation as such but rather the *marginal rate of transformation* between the good and bad output, we do not see any reason to restrict the shadow price to be only negative or zero. Indeed, in the pulp and paper industry, it is a well known fact that abatement measures may also increase pulp output. This was already noted in the pioneering work on water pollution by Kneese and Bower (1968) (see also Porter and van der Linde (1995 a,b). The basic reason behind this phenomenon is that the pulp plants have aimed to close the waste water system in order to decrease material wastes and consequently to increase productivity. Naturally, re-circulation of waste water simultaneously reduces bad outputs.

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, one can use a nonparametric or parametric model. For the present paper, the nonparametric model is excluded because it is piecewise linear and thus not differentiable and therefore the shadow price parameters cannot be determined directly. 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 specified. As is well known, the advantage of this form is its flexibility. Moreover, it does not impose strong disposability of outputs.

$$(3.1)\quad \ln D_o(x,v,w) = \alpha_0 + \alpha_1 \ln v_1 + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i + 1/2\,\alpha_{11}(\ln v_1)^2$$
$$+ 1/2 \sum_{i=1}^{M}\sum_{j=1}^{M} \beta_{ij}(\ln w_i)(\ln w_j) + 1/2 \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}(\ln x_i)(\ln x_j)$$
$$+ \sum_{j=1}^{M} \mu_{1j}(\ln v_1)(\ln w_j) + \sum_{j=1}^{n} \eta_{1j}(\ln v_1)(\ln x_j) + \sum_{i=1}^{M}\sum_{j=1}^{n} \xi_{ij}(\ln w_i)(\ln x_j).$$

In (3.1), v_1 denotes desirable (pulp) output, $w = (w_1,\ldots,w_M)$ undesirable (bad) outputs, and $x = (x_1,\ldots,x_n)$ inputs. The following symmetry (S) and homogeneity (H) restrictions are imposed:

$$(S)\quad \beta_{ij} = \beta_{ji}, \quad \gamma_{ij} = \gamma_{ji}, \quad \mu_{1j} = \mu_{j1}, \quad \eta_{1j} = \eta_{j1}, \quad \xi_{ij} = \xi_{ji}$$

$$(H)\quad \alpha_1 + \sum_{i=1}^{M} \beta_i = 1, \quad \alpha_{11} + \sum_{j=1}^{M} \beta_{ij} + \sum_{j=1}^{M} \mu_{1j} = 0, \quad \sum_{j=1}^{m} \eta_{1j} + \sum_{j=1}^{M} \xi_{ji} = 0.$$

The parameters of equation (3.1) are computed using the linear (or goal) programming formulation suggested by Aigner and Chu (1968). From the theory (see 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,

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

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

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

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" criterian to be the *minimum absolute error (MAE)* criterian, 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

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

$$s.t.$$

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

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

$$(iii)\quad \frac{\partial \ln D_o(x^k,v^k,w^k)}{\partial \ln v_1} \geq 0.$$

In addition to the above restrictions, the symmetry restrictions (S) and homogeneity restrictions (H) are imposed. Since the homogeneity constraint imposes proportionality on outputs, it 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 non-positive input derivatives (i.e. increasing inputs, with fixed level of output, cannot increase the value of the output distance function (efficiency)). Finally, constraint (iii) ensures that the desirable output shadow prices are greater than or equal to 0.

It may be noted that the computation of the above LP model differs from that of Färe et al. (1993) in two important 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, in terms of data, the present study uses panel data (152 observations), whereas Färe et al. (1993) use cross section data (30 observations). Consequently, the plant -and time -specific dummy variables are also included in the present LP model. The panel data LP model in this context is called by Lovell (1993) the "inter temporal goal programming approach." The model (3.4) is computed using the *GAMS* program.

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. (see e.g. Brown and Walker 1995). 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 Lovell et al. (1990) and Grosskopf et al. (1996) is used to estimate the stochastic distance function. This procedure imposes the value 1 on the distance function and uses the homogeneity property (i.e. the output distance function is homogeneous of degree +1 in outputs) to solve the invariance problem of the left-hand side of the distance function. 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 one of the bad outputs, i.e. the amount of suspended solids effluents (for details of the transformation, see pp. 34-35 of he study). This transformation imposes the homogeneity restriction and weak disposability of outputs. Experiments with other scaling variables (phosphorous and suspended solids effluents) indicated that the results are not sensitive to which variable is chosen for the transformation.

4. Empirical Results

The empirical analysis is based on data from the Finnish sulphate pulp industry, described in detail in Appendix II. The institutional background and the water pollution regulations concerning this sector are also discussed in detail in Appendix I. It suffices here to say that 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 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 of sulphate pulp output (GVP), net fixed capital stock (K), hours worked (L), value of materials input (M), biological oxygen demand (BOD), total waste water flow (FLOW), and suspended solids (SS) (see Appendix I).

Although, 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. In fact, all the other water pollution effluents are "subsets " of FLOW in the sense that the reduction of FLOW will also reduce all the other effluents. In the present data, the simple Pearson's correlation coefficient between the FLOW and BOD series is 0.48. Furthermore, when the logarithm of the BOD series was

regressed on the logarithm of FLOW and the second order of FLOW (period -and plant - specific dummy variables were also included), the results were as shown below (t-values shown in parentheses):

(4.1) LBOD = -3.50 + 2.59LFLOW - 0.31LFLOW2

(3.5) (3.9) (3.2) $\bar{R}^2 = 0.51.$

The regression results indicate that there is a statistically significant functional relationship between the BOD and FLOW variables. Consequently, we experiment also with a system model, in which the output distance function is estimated simultaneously with the equation for LBOD, the latter equation having the same specification as in equation (4.1).

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 for 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. This is an indication that the data does not allow proper identification of all the parameters of the second order terms. 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. 4.2), and (ii) a Cobb-Douglas form (eq. 4.3). The reason for including the second order and cross product output terms (rather than the second order and cross product terms of inputs) is that this allows for the analysis of joint outputs.

$$(4.3) \quad \ln D_o(x,v,w) = \alpha_0 + \alpha_p + \alpha_t + \sum_{i=1}^{m} \alpha_i \ln v_i + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i + 1/2 \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} (\ln v_i)(\ln v_j) + 1/2 \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} (\ln w_i w_j) + \sum_{i=1}^{m} \sum_{j=1}^{M} \mu_{ij} (\ln v_i)(\ln w_j),$$

$$(4.4) \quad \ln D_o(x,v,w) = \alpha_0 + \alpha_p + \alpha_t + \sum_{i=1}^{m} \alpha_i \ln v_i + \sum_{i=1}^{M} \beta_i \ln w_i + \sum_{i=1}^{n} \gamma_i \ln x_i ,$$

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 (FLOW)). Due to the panel nature of the data, equations 4.1 and 4.2 include plant-specific (α_p) and period- specific (α_t) fixed effects (see below). The plant and time subscripts are deleted for simplicity from eqs. (4.2-4.3); 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 (4.2) and (4.3) and the dependent variable and the output terms were transformed in order to impose homogeneity and allow for the estimation (see above). [3]

Since, a priori, there is usually not enough information to enable one to choose a stochastic specification which best approximates the *data generating mechanism*, a number of different specifications have to be estimated. In the present study, the stochastic model was estimated using the following specifications: a model without plant- and period-specific effects and four different panel data specifications (one- and two-way fixed and random effects models). To test whether plant- and period-specific effects should be included, and if so, whether they should be treated as fixed or random, we employed a likelihood ratio test (see Appendix.ESI.2). The test results indicated that the two-factor fixed effects (or covariance) model (TFFE) model should be used. However, the results showed that the fixed and random effects models 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 (P2-P8). Also, it is likely that plants are affected e.g. by 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 1.* show that in the stochastic models the measures of goodness of fit ($\bar{R}^2$) were high. The model diagnostics and specification test results are shown in Appendix ESI.1. Without going into the details of these diagnostics, it may be noted that the Hausman specification test statistic indicated that the right hand-side variables can be regarded as exogenous. Due to possible multicollinearity between the FLOW and BOD variables, the stochastic and linear programming models were also computed using only one of the bad outputs in time. Furthermore, the stochastic output distance functions were also estimated simultaneously with the model for BOD using the iterative *SURE* (seemingly unrelated regression equations) method. The results from all the different stochastic specifications indicated that the coefficient for the BOD variable is not statistically significant at the 5 % level. This was also true for the models in which only the BOD bad output was included, and in the models in which the output distance function was estimated simultaneously with the equation for BOD using SURE (I.e. models 3 and 10).

According to the different coefficient restriction tests (see Appendix ESI.1), the preferred model is the restricted translog model, in which only the FLOW bad output variable is included. Therefore, also the BOD shadow prices derived from these equations may not be very robust and should be viewed critically. However, because our interest is in analysing the linear programming and stochastic approaches in computing shadow prices, we have computed the shadow prices for all the different specifications and for both FLOW and BOD. These results will give an indication of how sensitive the shadow prices are to small changes in specifications and estimation methods.

*Table 1. Estimated parameters for alternative models (two-way fixed effects specifications and linear restrictions imposed; 152 obs.).**

Parameter	Linear Program Cobb-Douglas	Stochastic Cobb-Douglas	95% conf. min	95% conf. max	Linear Program Rest. Translog	Stochastic Rest. Translog	95% conf. min	95% conf. max
α_0	-0.506	0.627 (0.9)			-0.211	0.440 (0.5)		
γ_K	-0.188	-0.140 (3.6)	-0.217	-0.063	-0.196	-0.143 (3.1)	-0.223	-0.064
γ_L	-0.195	-0.157 (3.3)	-0.248	-0.065	-0.229	-0.153 (3.6)	-0.249	-0.057
γ_M	-0.495	-0.669 (12.1)	-0.777	-0.561	-0.484	-0.665 (11.7)	-0.777	-0.553
α_Q	0.896	0.881 (17.5)	0.783	0.979	0.825	0.934 (4.6)	0.531	1.337
β_{FLOW}	0.129	0.132 (2.7)	0.035	0.229	0.358	0.148 (0.9)	-0.162	0.458
β_{BOD}	-0.025	-0.013 (0.7)	-0.054	0.027	-0.183	-0.082 (0.9)	-0.268	0.104
α_{Q2}					0.116	-0.002 (0.0)	-0.051	0.046
μ_{QFLOW}					-0.203	-0.012 (0.8)	-0.042	0.018
μ_{QBOD}					0.087	0.014 (0.5)	-0.042	0.071
β_{FLOW2}					0.307	0.008 (0.3)	-0.056	0.072
$\beta_{FLOWBOD}$					-0.104	0.003 (0.1)	-0.060	0.067
β_{BOD2}					0.016	-0.018 (0.8)	-0.064	0.0285
$\bar{R}^2$		0.98				0.98		
Log-L		167.71				168.15		

** The White heteroskedasticity-consistent t-values are shown in parenthesis. Note: for simplicity, the fixed effects are not reported.*

In order to analyze the robustness of the parameters estimated using the econometric model we employed the bootstrap method. The method has been applied, among other purposes, to generate standard errors and confidence intervals for parameter estimates and elasticities (for a survey of the bootstrap applications in econometrics, see Vinod 1993). The bootstrap method derives standard errors for statistically derived quantities by means of Monte Carlo simulation. The model and parameter values are set at those estimated for the problem at hand, and the error distribution is taken to be the empirical distribution of the model residuals. A detailed description of the method used in the present study is provided in McCullough and Vinod (1993), pp. 3-4 (particularly the rescaled residuals approach is applied here). For each bootstrap trial (1000 used here), a set of artificial data is generated from the stochastic specification of the fitted model. The bootstrap estimates were computed using SHAZAM 7.0 program.

In *Table 2*, the first column reports the average of the bootstrap simulation values minus the observed values obtained from the original OLS estimates. The numbers in this column estimate the bias in the original parameter estimates. The results indicate that for both the Cobb-Douglas and restricted translog specifications the small sample bias is not a problem for the OLS parameter estimates. For most of the parameters there is difference only in the third decimal. The second and third column gives an indication how large is the difference between the asymptotic and bootstrap standard errors. The magnitudes of the two different estimates appear quite similar. In summary, the bootstrap results indicate that the parameter estimates are robust as regards the consistency of the estimates.

It should be noted that the bootstrap method could also be used in the case of the parametric linear programming model to generate standard errors for the parameters. Indeed, these estimates would provide an interesting basis for comparison between the stochastic and deterministic approaches. However, due to no readily available program for computing the bootstrap estimates for the parametric linear programming approach, we leave it for the future work to accomplish this.

*Table 2. Bootstrap Estimates for the Structural Standard Errors (1000 replications).**

Parameter	Cobb-Douglas means [a]	Bootstrap standard errors	Asymptotic standard errors [b]	Rest. Translog means [a]	Bootstrap standard errors	Asymptotic standard errors [b]
α_0	0.057	0.690	0.698	0.023	1.078	0.845
γ_K	0.000	0.040	0.039	-0.002	0.043	0.040
γ_L	-0.002	0.046	0.047	0.001	0.053	0.049
γ_M	-0.001	0.053	0.055	-0.002	0.059	0.057
α_Q	-0.001	0.051	0.050	-0.007	0.325	0.204
β_{FLOW}	0.001	0.048	0.049	-0.021	0.299	0.156
β_{BOD}	-0.001	0.021	0.020	-0.007	0.196	0.094
α_{Q2}				0.001	0.055	0.024
μ_{QFLOW}				-0.003	0.088	0.015
μ_{QBOD}				-0.002	0.063	0.029
β_{FLOW2}				-0.001	0.055	0.032
$\beta_{FLOWBOD}$				0.000	0.076	0.032
β_{BOD2}				0.000	0.024	0.023

**Note: for simplicity, the fixed effects are not reported. [a] The mean values are the bootstrap simulation values minus the "true" values obtained from the original data. The means are an estimate of the small bias. [b] Asymptotic standard errors are the standard errors of the original OLS estimates. They are shown here in order to make the comparison with the bootstrap estimates more readily available.*

4.1 Shadow Prices

The shadow prices for bad outputs computed using equation *(2.4)* are shown in *Table 3.* These were evaluated at the mean values of the data. For the stochastic models, the standard errors of the (nonlinear) shadow price (elasticity) parameters were also computed. Also, tests for the significance of the shadow prices in Cobb-Douglas specifications were run (for the restricted translog specification this is difficult to compute, due to the second order terms). Since the test is on a single non-linear hypothesis, the small sample distribution of the test statistic is unknown. However, the test statistic is asymptotically distributed as χ^2_q , where q is the number of hypotheses. The probability (*p*) values for testing the null hypothesis that the shadow price of FLOW is zero using the t -, F -, -and Wald tests were 0.010, 0.021, 0.019, respectively. Thus, we reject the null hypothesis at the 2 % significance level. However, for the shadow price of BOD, the respective *p* -values were 0.254, 0.507, 0.506; thus, we cannot reject the null hypothesis. Therefore, the use of the BOD shadow prices in the present case is mainly to compare the relative performance of the different approaches, rather than to draw policy conclusions.

Perhaps the most significant feature of the shadow price estimates, shown in *Table 3.*, is that the FLOW shadow prices are positive. Thus, marginal reductions of FLOW are associated with increase in revenues of the plants. In terms of *Figure 3.* in p.29 of this study, 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 increasing the re-circulation of waste water over time, the production of one ton of pulp in 1990 generated on average three times less waste water than in the 1972.

Table 3. Shadow Prices for Bads (FIM/ton), Evaluated at the Mean of the Data).

MODEL	FLOW	BOD
1. Cobb-Douglas LP	393.0	-76.4
2. Cobb-Douglas ST	406.7	-49.1
	(2.34)	(0.7)
3. Cobb-Douglas ST simultaneous	420.0	-54.3
4. Cobb-Douglas (only FLOW) LP	469.4	
5. Cobb-Douglas (only FLOW) ST	393.0	
6. Cobb-Douglas (only BOD) LP		90.1
7. Cobb-Douglas (only BOD) ST		-27.3
8. Rest. Translog LP	962.6	-432.9
9. Rest. Translog ST	318.8	-189.4
10. Rest. Translog ST simultaneous	348.0	-47.4
11. Rest. Translog (only FLOW) LP	962.6	
12. Rest. Translog (only FLOW) ST	396.4	
13. Rest. Translog (only BOD) LP		444.0

Note. LP and ST denote the *linear programming* and the *stochastic* models, respectively. For the stochastic Cobb-Douglas model, the t-values of the shadow prices are also shown, in parentheses.

The shadow price for BOD is negative in all the models except in the linear programming models 6 and 13. Thus, the results appear to indicate that the reduction of BOD is most likely achieved through diverting resources from production of pulp to abatement process (e.g. building waste water treatment plants). The difference in the shadow prices between FLOW and BOD probably reflects the fact that the non-regulated bad output (FLOW) and regulated bad output (BOD) have been reduced to some extent by using different measures. FLOW has primarily been reduced by *pollution-prevention methods* (re-circulation), whereas BOD has been reduced mainly using the *end-of-pipe measures*, such as building biological (aerated) treatment lagoons.

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 pioneering study of water pollution, Kneese and Bower (1968) noted that the reduction of waste water from pulp and paper mills is to some degree the result of the plants objective to increase productivity. Thus, waste water reduction may have emerged as a by-product of productivity improvement measures rather than due to regulation as such. However, without doubt the regulation has also forced the plants to adjust production process and reduce pollution, but it is difficult to separate the impact of these two factors on the waste water reduction.

It is interesting to compare the present results to those obtained by Färe et al. (1993), who use a linear programming output distance function and cross-section data to derive shadow prices of water pollution effluents for pulp and paper mills operating in Michigan and Wisconsin in 1976. The results obtained by Färe et al. showed that on average reducing one ton of BOD emissions diverts enough resources to have produced over two tons of paper (the Färe et al. 1993 study did not include FLOW effluents). This is a much greater impact than the one indicated by the present study. Depending on the model specification in *Table 3*, the shadow prices indicate that a ton reduction of BOD is associated at worst (model 8) with only a 16 % reduction in pulp output, whereas in the most favourable case (model 13) the reduction of BOD by one ton would be associated with an 16% increase in pulp output. 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 different production technologies or/and that they are operating in a 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 the shadow prices to be non-positive). 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.

The results from the present study also indicate that one should be careful in restricting the shadow prices of undesirable outputs to be non-positive. Usually, this restriction is imposed on the grounds that the bads are subject to regulation and that there must be a trade-off of reducing these in terms of lost "good" output (see e.g. Färe et al. 1993 and Coggins and Swinton 1996). However, as stated above, the shadow price does not reflect the impact of regulation as such but rather the marginal rate of transformation between the bad and good output. If bads are an indication of inefficiency of the production process, the reduction of these can in fact increase the revenues of the firm. At least in the pulp and paper industry, there is plenty of evidence that reduction of water pollution may also increase production efficiency (e.g. Kneese and Bower 1968, Gray and Shadbegian 1993, Porter and van der Linde 1995 b).

4.2 Comparing the Approaches

How do the estimates using the two different methods compare? Analysing the results shown in *Table 2* shows that with the stochastic approach, the Cobb-Douglas and restricted translog functional specifications lead to essentially the same conclusions concerning shadow prices, but with the linear programming approach, the two functional forms give rather different results. For the stochastic models, the mean values (over the data) of the FLOW shadow prices are very close and the BOD shadow prices have the same sign. In contrast, with the linear programming approach, depending on the functional form, the shadow prices for FLOW vary greatly and for BOD even change the sign. In general, the experiments with the linear programming approach, in which either the functional form or data were changed slightly, indicated that minor changes in the specification led to major changes in the inferences.

It may be noted that the results for the FLOW shadow prices are not very sensitive to whether the stochastic output distance function is estimated simultaneously with the model

for BOD using SURE or without this model using OLS. The only notable change in shadow prices computed with these two different specifications is in the BOD shadow price in the restricted translog specification (see models 9 and 10).

Since the model diagnostics for the stochastic specifications indicated that the restricted translog model with only the FLOW bad output included performed the best, it is interesting to examine more closely the differences between the two approaches for this specification. The *Box and Whisker* plot of the FLOW shadow prices computed over the data are shown in *Figure 1*. The plot is used as a graphical method for comparing the distribution of the shadow prices computed using the two approaches. In the plot, the mean of the shadow price is indicated by a small box in the center, while the dispersion (variability) is represented by ±1 times the standard deviation. The interval demarcated by the "whiskers" indicates the values which fall in the range of the mean ±1.96 times the standard deviation. Thus, for normally distributed variables the range of the whiskers will include approximately 95 % of the observations. For the stochastic model the mean FLOW shadow price is 319 (FIM) and 95 % of the values fall between 220 - 420 (FIM). The corresponding mean value for the linear programming model is 963 (FIM), and 95 % of the values fall in the range 400 - 1 500 (FIM). Thus, the linear programming approach gives a mean value which is over three times as large as that obtained with the stochastic model. Furthermore, in *Figure 2*, which shows the same series scaled by the respective mean values, the variation (standard deviation) is larger for the linear programming model than for the stochastic model.

Although the level of FLOW shadow price generated by the stochastic and linear programming approaches differ, the changes over time in the shadow prices are very similar. *Figure 3*, shows the mean FLOW shadow prices for each year between 1972-90. With both approaches, the shadow price follows a similar downward trend. In summary, the results of the study indicate that the linear programming shadow prices are sensitive to specification of functional form, whereas the stochastic approach produces rather similar results for all the different specifications applied in the analysis.

Figure 1 . Box & Whisker Plot of FLOW Shadow Prices (FIM/ton)

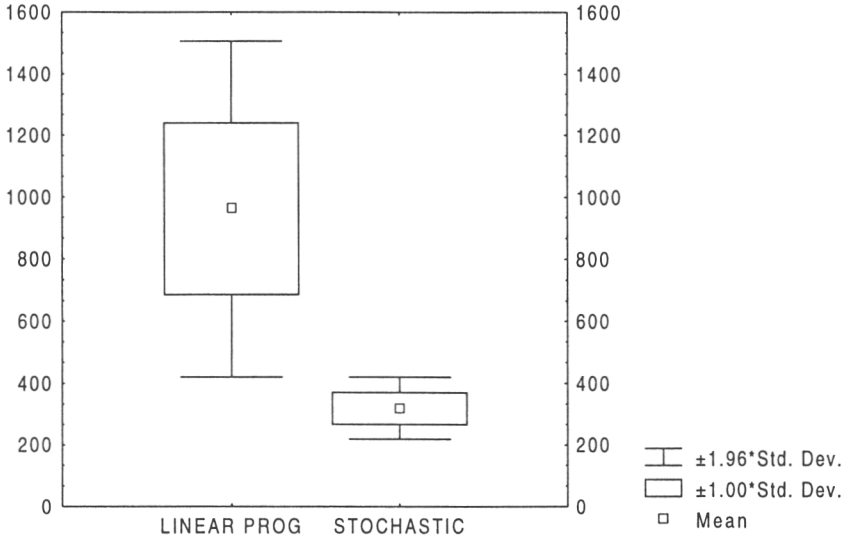


Figure 2 . Box & Whisker Plot of FLOW Shadow Prices (scaled by mean values)

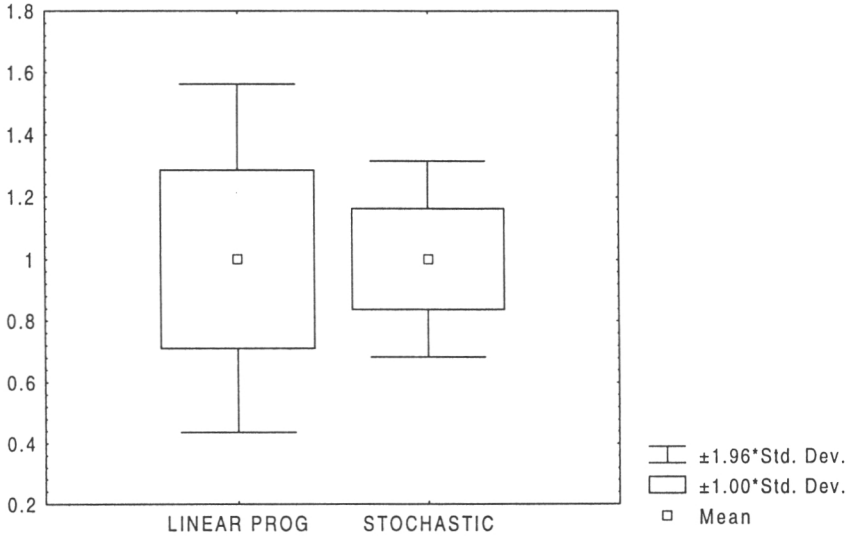
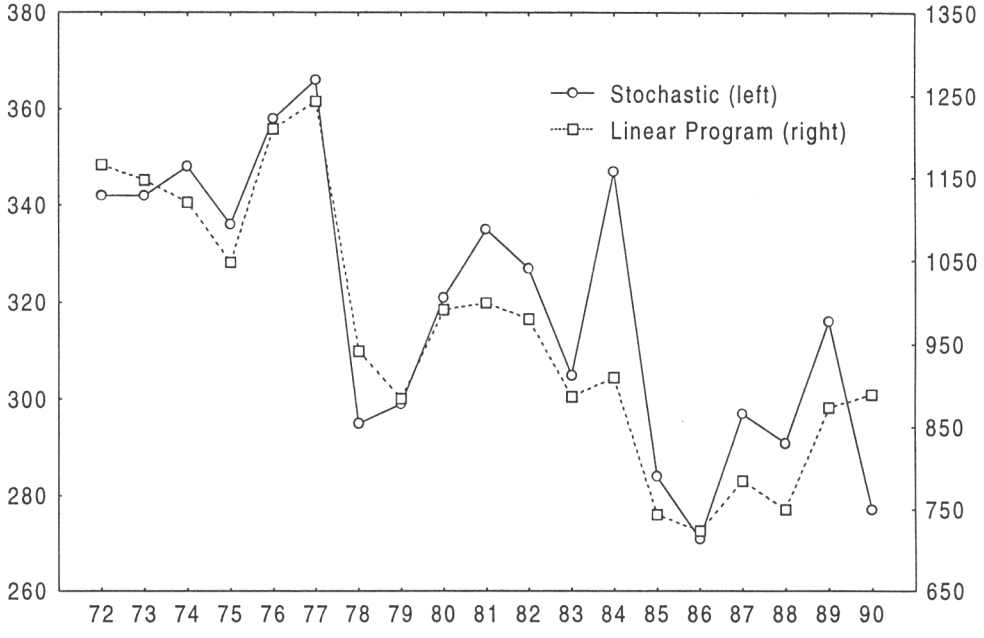


Figure 3. FLOW Absolute Shadow Prices (FIM/ton) (evaluated at the mean values of data, 1972-90)



Finally, there is an important question which arises with the use of the linear programming methodology. Namely, if only the linear programming approach had been used (as in most of the previous studies), which model specification would we have ended up with? Probably it would have been 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, because of which the parameters could not be precisely identified (and the stochastic model not even estimated). Moreover, how could we have decided whether to include in the model the BOD bad output or the plant and time -specific fixed effects? These questions underline the important drawback with the parametric linear programming approach, i.e. the lack of tools to guide the model specification.

5. Summary and Implications

In the present study the econometric and parametric linear programming techniques were used to compute shadow prices for waste water effluents in the pulp industry. The implications of the results are of two kinds: those related to the effects of pollution control on sulphate pulp plants and those related to methodological issues. The substantive implications are summarized first.

The results of the two approaches used are coherent in that both indicate that the reduction of waste water flow by Finnish pulp plants has been associated with the increase in revenues. However, the positive FLOW shadow prices should not be interpreted as showing 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. Another is the apparent strong positive correlation between pollution control measures and improvements in the production process. The result also indicates that one should not a priori rule out the possibility that pollution control may be positively correlated with increases in firms' revenues. In contrast, the shadow price for BOD was in most specifications negative; thus, there appears to be a trade-off in reducing BOD in terms of foregone pulp output. However, the specification tests for the stochastic models indicated that the BOD variable does not contribute significantly to the explanatory power of the models, and thus the results for BOD are probably not very robust.

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 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 of using 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 in the deterministic approach.

The results of this study confirm the existence of these problems for the deterministic approach. For the data used in the present study, the results for the deterministic model were more 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 tested and the sensitivity of the results to small changes in the data should be examined.

Notes

1. 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.
2. 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 $v \leq y \in P(x)$ implies $v \in P(x)$. 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.
3. It may be noted that there is no study in which factor share equations are derived from the output distance function. Thus, efficient system estimations, such as for the cost function and respective factor share equations, cannot be computed for the output distance function.

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.

Brown, B. W. and M. B. Walker (1995). Stochastic specification in random production models of cost-minimizing firms, *Journal of Econometrics*; 175-205.

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

Coggins, J. S. and J. R. Swinton (1996). The Price of Pollution: A Dual Approach to Valuing SO2 Allowances, *Journal of Environmental Economics and Management*, 58-72.

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

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.

Ferrier, G. D. and C. A. K. Lovell (1990). Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence, *Journal of Econometrics*, 46; 229-245.

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

Gong, B. and R. C. Sickles (1992). Finite Sample Evidence on Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data, *Journal of Econometrics*; 259-284.

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

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

Grosskopf, S., D. Margaritis, and V. Valdmanis (1995). Estimating output substitutability of hospital services: A distance function approach, *European Journal of Operational Research* 80; 575-587.

Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1996). Budget Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling, *Review of Economics and Statistics*, forthcoming.

Gray, W. B. and R.J. Shadbegian (1993). Environmental Regulation and Manufacturing Productivity at the Plant Level, National Bureau of Economic Research Working Paper No. 4321.

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.

Kneese, A.V. and B.T. Bower (1968). *Managing Water Quality: Economics, Technology, Institutions*. Resources for the Future, John Hopkins Press.

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.).

Porter, M. E. and C. van der Linde (1995a). Toward a New Conception of the Environment-Competitiveness Relationship, *Journal of Economic Perspectives*, vol 9, No. 4; 97-118.

Porter, M. E. and C. van der Linde (1995b). Green and Competitive: Breaking the Stalemate, *Harvard Business Review*, September-October; 120-134.

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

Appendix ESI. 1. Diagnostics and Specification Tests *

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) is accepted for 6 plants out of the total of 8 plants.

Heteroskedasticity. Whites's heteroskedasticity test, based on F-statistic was used. The computed F-statistic for Cobb-Douglas specification is F = 1.38, and for the restricted translog F = 1.53. Thus, homoskedasticty cannot be rejected at the 5% significance level. However, the White heteroskedasticity-consistent standard errors are shown for the parameters.

Normality. The *Jarque-Bera (JB) -statistic* was used to test whether the residuals are normally distributed. The null hypothesis of normality was rejected at the 5% significance level, both for Cobb-Douglas (*JB = 22.45*) and restricted translog (*JB = 24.32*) specifications. However, when 4 outlier observations (obs. 87,118,119,120) were removed, the null hypothesis of normality was accepted (*JB = 1.23* and *JB = 1.71* for Cobb-Douglas and restricted translog, respectively). It may be noted that the results for the shadow prices were not sensitive to estimating the models without the above four observations.

Orthogonality. The *Hausman specification statistic (H)* was used to test the exogeneity of the output terms on the right-hand side of the regression equation. The test indicated that (at the 5 % significance level and with three degrees of freedom) OLS is an consistent estimation method, in contrast to the instrumental variables estimator ($H_3 = 7.15$).

Testing parameter restrictions. The *Wald test* of coefficient restrictions was carried out to examine which of the functional forms should be preferred and to examine whether the undesirable outputs (FLOW and BOD) should be included in the models. First, in the Cobb-Douglas specification, the F-statistic for testing that the coefficient of the BOD variable is zero, produced an F = 0.42. Thus, we would accept the null hypothesis, even at the 1 % significance level. A similar test for the coefficient of the FLOW variable produced an F = 11.28; thus we would reject the null even at the 1% significance level.

A test on the Cobb-Douglas vs. restricted translog form produced an F = 0.22, thus indicating that one should accept the Cobb-Douglas in favor of the restricted translog, even at the 1 % significance level. However, once the BOD variable was omitted from the specifications, the computed F was 7.20. Consequently, the restricted translog form with the BOD variable omitted is the preferred specification.

Functional form specification. The Ramsey RESET statistic was used to test the null of correct functional form specification of the original model against the alternative that the squared fitted values have been omitted. In the Cobb-Douglas and restricted translog specifications the coefficient for the squared fitted term had a t-values of 0.5 and 1.75, respectively, thus giving support to the original functional forms.

**See Appendix IV Test Statistic for detailed description of the tests*

Appendix ESI. 2. Test Statistics for Panel Data Model Selection

COBB-DOUGLAS

Model	LR-test	Chi-sq.	d.f.	Decision
1. constant term only	2 vs 1	126.48	7	choose 2
2. plant and period fixed effects only	3 vs 1	468.49	6	choose 3
3. X -variables only	4 vs 1	500.21	13	choose 4
4. X and plant fixed effects	4 vs 2	373.72	6	choose 4
5. X and plant and period fixed effects	4 vs 3	31.72	7	choose 4
6. X and plant and period random effects	5 vs 4	143.74	18	choose 5
	5 vs 3	175.45	26	choose 5

RESTRICTED TRANSLOG

Model	LR-test	Chi-sq.	d.f.	Decision
1. constant term only	2 vs 1	126.48	7	choose 2
2. plant and period fixed effects only	3 vs 1	481.44	12	choose 3
3. X -variables only	4 vs 1	519.80	19	choose 4
4. X and plant fixed effects	4 vs 2	393.32	12	choose 4
5. X and plant and period fixed effects	4 vs 3	38.36	7	choose 4
6. X and plant and period random effects	5 vs 4	143.73	18	choose 5
	5 vs 3	182.09	26	choose 5
	5 vs 6	14.84	12	choose 5

ESSAY II

Estimating Shadow Prices for Bads Using a Two-Stage Stochastic Distance Function Model

Lauri Hetemäki

Abstract

Pollution control usually requires firms to adjust their production processes. The purpose of the present study is to examine to what extent these adjustments affect firms' revenues and performance in general. The theoretical framework is based on the output distance function, which is used to derive the shadow prices of pollution and the impact of pollution reduction on firms' performance. The present study differs from the previous literature in that it uses a stochastic distance function rather than a deterministic one, by employing plant level panel data and imposing no a priori restrictions on the values of the shadow prices of pollution. Also, a novel two-stage approach which combines deterministic nonparametric linear programming with a stochastic econometric model is used to estimate the distance function. These extensions enable the estimation of a more general model, whose statistical significance can be explicitly assessed. The method is used to analyze the effects of water pollution regulations on eight pulp plants in Finland observed over a period of 19 years (1972-90).

I would like to thank Shawna Grosskopf, Mika Hakuni, Pekka Ilmakunnas, Jari Kuuluvainen, C.A. Knox Lovell and Markku Rahiala for their comments on a earlier version of this paper. Also helpful comments were received at the TIMS XXXII Conference, Anchorage, Alaska, June 12-15, 1994, and Nordic Workshop on Productivity, Growth and Development, Göteborg, January 26-28, 1994. This research was supported by the Academy of Finland and the Nordic Council of Ministers Project "Economic Instruments in Environmental Policy: The Nordic Forest Sector". A previous version of this paper appeared in the Finnish Forest Research Institute, Research Reports series under the title "Do Environmental Regulations Lead Firms Into Trouble? Evidence From a Two-Stage Distance Function Model with Panel Data".

1. Introduction

In the last decade or so, there has been a surge of empirical literature which attempts to explicitly incorporate and assess the impacts of environmental regulations on producer behavior (e.g., Barbera & McConnell 1990, Conrad & Morrison 1989, Gollop & Roberts 1983, Jorgenson & Wilcoxen 1990, Myers & Nakamura 1980, Pittman 1981). This is a natural outcome of the increasing importance of environmental regulations and of the fact that pollution (undesirable output) can no longer be disposed of without cost. Firms have to adjust their production processes in order to reduce pollution to permitted levels or pay additional charges for their effluents. In the literature, interest has centred on the assessment of the effects of the regulations on productivity, efficiency, economic growth and factor demand, among other issues. The "standard approach" for including the environmental aspect in this context has been to add an additional input, such as abatement capital, to the producer's factor demand function or to introduce an environmental tax parameter for some input or output. Therefore, the inclusion of regulation or effluents has not essentially changed the conventional empirical neo-classical models of producer behavior. The purpose of this study is to use a new approach, namely, the output distance function framework, to analyze the effects of pollution control on Finnish pulp and paper plants.

Shephard (1970, 1974) noted that the conventional assumptions and models of producer behavior should be modified in the case where the production process generates undesirable outputs which cannot be freely disposed.[1] Indeed, for many production processes which generate regulated undesirable outputs, the conventional single-product firm framework is not appropriate. Pollution (undesirable output) is often a side product created as a result of the production of the "good" (desirable) output. In other words, pollution does not necessarily enter the production process as an input but is instead created jointly during the production process. Consequently, there is generally a trade-off (marginal rate of transformation) in production between the desirable and undesirable outputs, the quality and quantity of which is of central importance when assessing the impact of pollution control. According to Shephard (1970), the polluting firm's production process is more accurately modeled as a production of multiple outputs, with strong disposable

desirable outputs and weak disposable undesirable outputs. Färe et al. (1993) have extended Shephard's framework in order to examine the effects of waste water regulations on U.S. pulp mills. They use the output distance function as an analytical tool for representing the production technology and derive shadow prices for undesirable outputs.

Besides addressing the above issue concerning the way pollution is modeled in the production process, the empirical literature to date has often been based on somewhat unsatisfactory data. Due to a lack of micro level data, these studies have generally concentrated on measuring the effects of environmental regulations or pollution control using industry or country level aggregate data. However, if pollution regulations are set, for example, at the plant level, it is also important to be able to measure the effects of pollution control at the plant level. Firm or plant level regulations are particularly common for point source waste water effluents. For example, in the pulp and paper industry in Finland and Sweden, every plant is individually regulated with respect to the substances that it discharges. Similar practices are also common in many industries in the United States and Canada. Besides the "conventional" aggregation bias, using industry level data in such cases causes identification problems and measurement errors.

The purpose of the present study is 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 using the theoretical approach of Shephard (1970) and Färe et al. (1993). However, while Färe et al. use deterministic linear programming analysis and plant level cross section data, the present study is based on the *stochastic* distance function and plant level *panel data*. Moreover, Färe et al. restrict the shadow prices of undesirable outputs to be nonpositive, whereas no such a priori restriction is set in the present study.

Finally, a new two-stage estimation procedure is formulated for the stochastic distance function. The method differs from the existing stochastic distance function studies in that it uses the actual data dependent distance measures to estimate the distance function. In line with the previous studies, the present study also utilizes the property that the output distance function is homogeneous of degree +1 in outputs to transform the equation to an estimable form. The previous studies are a special case of the present two-stage method. In particular, in a case, in which for all the observations the distance scores are equal to one, the present method reduces to the one used by the previous studies. In addition, although the

nonparametric distance scores are used primarily for the purpose of the estimation procedure, they also provide information on the technical efficiency of plant production.

2. Methodological Background

Although the assessment of pollution control costs is straightforward in principle, in practice the problem is that the regulatory agency rarely knows the marginal abatement costs of individual plants. In general, reliable data on abatement costs and the resulting reduction in effluents are not available. Thus, one usually must try to infer the marginal treatment costs indirectly. As Färe et al. (1993) show, one possible way to derive the marginal treatment costs indirectly is to use the duality of the *output distance function* and the revenue function. This approach also provides information on the production technology of the plants. Furthermore, Färe et al. assume *weak disposability* of undesirable (bad) outputs, which allows them to model the fact that regulations restrict the firms' ability to costlessly dispose of effluents. The output distance function allows the computation of shadow (virtual) prices of pollutants without requiring detailed information about the actual regulations or abatement costs. Furthermore, the method allows one to identify the shadow prices of pollution at the level of the individual plant. These shadow prices reflect the impact of pollution control on a plant and indicate to what extent the revenues of the plants are affected by pollution control measures.

The output distance function has important advantages over more traditional means of representing production technology. In comparison to a production function, a distance function allows one to model multiple output and joint production technologies. On the other hand, the advantage of the distance function over cost, profit and revenue functions (which can also be readily used to model multiple output technology) is that no maintained behavioral hypothesis (cost minimum or profit/revenue maximum) is required. A distance function only identifies the technology frontier and gives the distance to the frontier for each observation. As a result, the different measures of economic effects (e.g. substitution) are not conditional on the behavioral hypothesis. Also of great practical importance is the property that the distance function can be computed with data on quantities of inputs and outputs alone; prices are not needed.

In empirical applications of distance functions, it has been common practice to use deterministic linear programming; the econometric approach has rarely been used. Indeed, Lovell et al. (1990, footnote 6) state that *"Although empirical computation of distance functions using econometric techniques is certainly in its infancy, mathematical programming techniques have been used to calculate distance functions for many years now."*

As far as we know, the only empirical applications of the parametric output distance function in the environmental regulations literature are Färe et al. (1993), Hetemäki (1994, 1995), Coggins and Swinton (1996), and a mimeo by Brännlund (1996). The studies by Färe et al. and Coggins and Swinton use the deterministic parametric linear programming approach. The problem with the deterministic model is that the computed parameters may be affected by random factors not controlled by the plants, which may be numerous and complex and not observable, and hence not measurable. On the other hand, the stochastic model allows for random error and also permits direct testing for statistical significance and consistency of the estimates. Finally, there may be production processes for which it may be inappropriate to restrict the shadow prices of undesirable outputs to be nonpositive, as Färe et al. (1993) and Coggins and Swinton (1996) do.[2] In any case, because the axioms behind the theoretical model do not require such a restriction, it is unecessary to impose it. Consequently, in the present study the shadow prices of "bad outputs" are not restricted to be nonpositive. Finally, using panel data rather than cross section data, provides more precise and consistent estimation even in the presence of correlated plant-specific effects (Hsiao 1986). Naturally, panel data also allow one to examine the impacts of pollution control over time.

The approach used here differs from that of the existing stochastic distance function studies in that the empirical analysis is based on a two-stage approach. In the first stage, a nonparametric linear programming model is used to compute the distance measures (efficiency scores) for each plant. In the second stage, these distance measures are used as a dependent variable in the stochastic output distance function model. The main advantage of this approach is that the assumption that the producers are operating on the production frontier (i.e. the value of the distance function is set equal to 1) can be relaxed. This provides a more realistic approximation of the production technology for most applications.

3. Theoretical Model

The initial incidence of much of pulp industry pollution control falls on the firms. The ultimate incidence depends on the ability of firms to shift the cost burden to consumers by raising prices or to workers or other factors of production. The theoretical model in this study focuses on the initial incidence of pollution control and therefore is restricted to a partial equilibrium analysis of how a firm (or plant) reacts to a change in its cost structure. Within this framework, the output distance function can describe the effects of pollution control on the production technology of a firm. The conventional production function gives the maximum output obtainable from a given input vector. The distance function is a generalization of this notion, serving as a functional representation of the output set in the context of multi-outputs (Shephard 1970, Färe 1988).

A production technology transforming factors of production $x = (x_1, x_{2,\ldots,} x_n) \in R_+^n$ into outputs $y = (y_1, y_2 \ldots, y_m) \in R_+^m$ can be modeled by the output set *P(x)*. The output 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 maintained axioms of Färe (1988, p.6). In particular, outputs are assumed to be only *weakly disposable*. Conventionally, *strong* (or free) *disposability* is assumed.[3] However, for the pulp industry, it is unlikely that the production of pulp is characterized by strongly disposable outputs, given that regulations do not allow water pollution to be "thrown away" (freely disposed). The output distance function is defined on the output set *P(x)* as

$$(3.1) \quad D_o(x,y) = \min_{\theta} \{\theta : (y / \theta) \in P(x)\}.$$

Equation *(3.1)* 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 (for a detailed description, see Färe 1988, pp. 31-34). The value of the output distance function must be less than or equal to one ($D_o \leq 1$) for feasible output. Further, the value of the distance

function is the reciprocal of the *Farrell output-based technical efficiency index* (Färe 1988). The revenue function defined as (Shephard 1970 , Färe 1988)

$$(3.2) \quad R(x,r) = \max_{y}[ry: y \in P(x)],$$

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,

$$(3.3a) \quad R(x,r) = \max_{y}\{ry: D_o(x,y) \leq 1\}$$

$$(3.3b) \quad D_o(x,y) = \max_{r}\{ry: R(x,r) \leq 1\}.$$

The duality theorem shows that 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.

According to Shephard (1970) and Färe et al. (1993), the *absolute* output shadow prices (r_m^s) for each observation of undesirable outputs ($m = 2,....,M$) can be derived as the derivative of the distance function (using dual Shephard's lemma), i.e.

$$(3.4) \quad r_m^s = R^*(x,r^s) \bullet (\partial D_o(x,y) / \partial y_m) = r_1^o \bullet \frac{\partial D_o(x,y) / \partial y_m}{\partial D_o(x,y) / \partial y_1}.$$

In equation *(3.4)* the ratio of output shadow prices reflects the relative opportunity cost of the outputs, i.e. they are 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 we often do not have data on regulations, and even when such data

exists, the plants rarely operate exactly at the level of the constraint. Further, the shadow prices do not require the plants to operate on the production frontier. Since the output distance function is homogeneous of degree +1 in outputs, the derivatives, which give the shadow prices, are homogeneous of degree zero with respect to proportional scaling of outputs. Because the output distance function is such a proportional scaling of outputs, the shadow prices are independent of whether the observations are on the frontier (see Grosskopf and Hayes 1993).[4]

4. Empirical Model

Let the production technology of the pulp plants be represented by the output distance function

$$(4.1) \quad D_O = f(X, Y; \Psi) \exp \varepsilon ,$$

where D_0 is the distance measure, $f(.)$ is the production technology, X is a matrix of inputs, Y is a matrix of outputs, Ψ is a vector of parameters to be estimated and ε is the error term. Estimation of a distance function like *(4.1)* raises several econometric problems, and presumably because of these, there have been very few econometric applications of distance functions. The basic problem with distance functions, as concerns econometric estimation, is that one does not 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. $D_O = ln(1) = 0$).

In order to avoid the above problems, Lovell et al. (1990) and Grosskopf et al. (1996) utilize the property that the output distance function is homogeneous of degree +1 in outputs (see also Grosskopf and Hayes 1993). Thus, for each observation to be used in estimating the distance function, a value that is unique to that observation can be used to multiply all output values on the right hand-side and the value of the distance function on the left-hand

side. However, this transformation may cause estimation problems. After the transformation, the multiplicative variable appears on both the left-hand and right-hand sides of the equations, which may result in endogeneity on the right-hand side. Thus, one has to test whether the errors are correlated with the regressors and if so, instrumental variables estimation must be used.

Another problem in estimating the output distance function is that in theory the value of the distance function should never exceed that for a plant operating on its frontier. However, in the estimation of equation *(4.1)* an error term with mean zero but positive variance is assumed. For some plants the forecasted value of the output distance function can therefore exceed the theoretically plausible value. To account for this problem, it is common (e.g. Lovell et al. 1990, Grosskopf et al. 1996) to use the method known as corrected ordinary least squares (COLS). This amounts to calculating first the most negative residual from the estimated output distance function and then adding that residual to the intercept term so that the corrected estimate of the output distance function never exceeds the theoretically plausible value for any plant. In other words, this ensures that all observations are enveloped from above.

In the present study, a two-stage approach is used to estimate the output distance function. In short, the procedure consists of two steps: first, the measure for each plant's distance (in each period) to the reference production frontier is computed using a deterministic nonparametric piecewise linear model, which treats desirable and undesirable outputs differently. Second, the distance measures computed in the first stage are used as a dependent variable in a parametric stochastic distance function model. By using this approach, one can relax the assumption that all plants are operating on the frontier. Existing studies have either assumed that all observations have an equal distance measure of 1, or inefficiency has been introduced afterwards by means of a composed error structure, as in Grosskopf and Hayes (1993). Naturally, the larger the dispersion of the distance measures from 1, the more biased the results obtained using methods that set the distance scores equal to 1. Moreover, compared to the "conventional approach" of estimating a stochastic distance function, the two-stage approach will most likely have an impact both on the intercept and the slope parameters of the regression equation. This is because the dependent variable is likely to have a different variance in the two-stage method as compared to the conventional approach.

Thus, in order not to restrict the technology to being on the frontier, we compute the so-called *Farrel output efficiency measure*, *F*, which has been shown to be the reciprocal of the output distance function (e.g. Färe 1988, p. 135). Consequently, the dependent variable of the stochastic output distance function can be computed in a theoretically consistent way. From a slightly different perspective, the two-stage procedure consists of first constructing the production frontier using a nonparametric linear programming model and then approximating this frontier with a smooth, parametric functional form, which provides *additional and economically and statistically interpretable results*. The advantages of complementary use of nonparametric linear programming and stochastic econometric models have also been noted by Banker and Cooper (1994).

It may be noted that recently a similar type of two-stage approach has been applied in the Data Envelopment Analysis (DEA) literature (see e.g. Grosskopf 1996). However, the two-stage approach used in the present study differs from this literature in that the interest is not to analyze and explain the efficiency scores. In the DEA literature the second stage econometric models are always *ad hoc*, including variables which are considered not to be directly controllable by the decision making units (DMU), but that are important in explaining the variations in the efficiency scores (see Grosskopf 1996). In the present study the second stage model is based on econometric output distance function, and the interest is to analyze the shadow prices of bad outputs, not the efficiency.

For the model specification with strongly disposable desirable outputs, weakly disposable undesirable outputs and variable returns to scale (VRS) (with respect to inputs), *F* can be expressed as (for a more detailed description, see Färe et al. 1994, pp. 105-106)

$$(4.2) \quad F^k_{vrs}(v^k, w^k, x^k) = \max \pi$$

$$(i) \quad \pi v^k \leq \phi z^1 v^1 + \phi z^2 v^2 + \ldots + \phi z^k v^k$$

$$(ii) \quad \pi w^k = \phi z^1 w^1 + \phi z^2 w^2 + \ldots + \phi z^k w^k$$

$$(iii) \quad x^k \geq z^1 x^1 + z^2 x^2 + \ldots + z^k x^k$$

$$(iv) \quad \sum_{k=1}^{K} z^k = 1, z^k \geq 0, \forall k \quad , 0 \leq \phi \leq 1,$$

where v and w denote the desirable and undesirable output subvectors of y respectively, and z the intensity or scaling vector, and k ($k=1,....,K$) denotes the observation. The z vector enables one to shrink or expand individual observed activities for purpose of constructing unobserved but nonetheless feasible activities. In other words, z provides weights which facilitate the construction of the linear segments of the piecewise linear boundary of the technology.

Constraint (i) says that the reference unit must produce at least as much good output as unit k ; constraint (ii) imposes weak disposability of bad outputs and states that the reference unit must produce the same amount of bad output as unit k; constraint (iii) states that use of inputs of unit k must at least equal the amounts employed by the reference unit; and constraint (iv) restricts the technology to permit variable returns to scale (the constant returns to scale (CRS) specification is obtained by setting the restriction (iv) as $\sum_{k=1}^{K} z^k \geq 0$).

Problem *(4.2)* can be linearized by setting $\phi = 1$. This procedure does not affect the maximizing π, z values. Model *(4.2)* is a very general model, entailing many different models as special cases. Furthermore, the computed distance measures are not dependent on any particular functional form (except piecewise linearity). The F index is computed by comparing each input-output combination (x,y) to a reference technology set formed from all observations. The frontier consists of piecewise linear facets, which are determined by the efficient units of the data. In accordance with the output distance function, the reciprocal of F gets values $0 \leq F \leq 1$.

The potential weakness of the two-stage approach is that it may cause endogeneity of the right-hand side variables. Since the F measure is computed using the same quantity and input data that are used as exogenous variables in the stochastic model, the explanatory variables may be correlated with the equation error and the least squares estimates of the coefficients may be biased and inconsistent. Consequently, one should check on whether endogeneity is a problem and if so, use the instrumental variables method to estimate the model.

In order to be able to estimate the parametric stochastic distance function, a functional form has to be chosen. In principle, it would be desirable to use as flexible a functional form as possible. Initially, a translog form including 28 parameters was tried. However, the

translog model could not be estimated consistently due to multicollinearity. Consequently, the estimated output distance function is a special case of a translog function, with a first-order approximation in the input quantities (Cobb-Douglas technology) and second-order terms in the output quantities; it is shown as eq. *(4.3)* (the same functional specification was used e.g. by Simar 1992).

$$(4.3) \quad \ln D_{opt} = \alpha_0 + \sum_{n=1}^{N} \beta_n \ln x_{npt} + \sum_{m=1}^{M} \gamma_m y_{mpt} + 1/2 \sum_{m=1}^{M} \sum_{m'=1}^{M} \gamma_{mm'} (\ln y_m)(\ln y_{m'}) + \varepsilon_{pt},$$

where x_n denotes inputs, i.e. capital (K), labor (L) and materials (MA); y_m denotes desirable and undesirable outputs, i.e. quantity of pulp production (Q), biological oxygen demand (BOD) and waste water flow (FLOW); p indexes the plants ($p = 1,2,..,P$); and t denotes the time period ($t = 1,2,...T$). For the estimation, the homogeneity restriction (i) and the symmetry restriction (ii) are set as

$$(i) \quad \sum_{m=1}^{M} \gamma_m = 1, \quad \sum_{m'=1}^{M} \gamma_{mm'} = 0, \quad m = 1,...,M$$

$$(ii) \quad \gamma_{mm'} = \gamma_{m'm}, \quad m = 1,...M, \; m' = 1....M$$

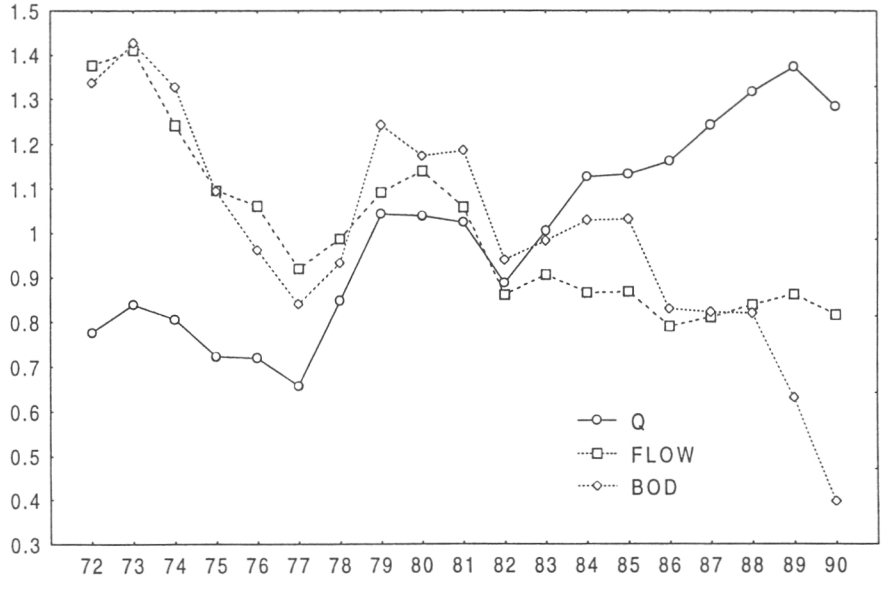
5. Results

In order to keep the sample as homogeneous as possible, and thus reduce the bias of comparing plants with different vintages of production technology, only those plants were included which were operating during the whole period studied. Moreover, the linear programming model requires a balanced panel. Consequently, the data sample contains annual data from 8 sulphate pulp plants observed over the period 1972-90 (a detailed description of the data is given in Appendix II). All the plants are nonintegrated, 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 about one-half of the total production of the sulphate pulp industry during

1972-90. The pulp output and the effluents of the plants over time are shown in *Figure 1.* (the data have been divided by their respective mean values). The figure shows that over time there has been a simultaneous increase in pulp output and decrease in effluents.

The data used for estimation consists of observations on quantity (Q) of sulphate pulp output, net fixed capital stock (K), hours worked (L), value of materials input (M), biological oxygen demand (BOD), and total waste water flow (FLOW). Although FLOW has not been regulated by the water authority, its reduction has nevertheless been one of the major means by which the plants have tried to reduce different water pollution substances. In particular, the reduction of FLOW describes the internal process changes adopted in order to reduce waste water. Moreover, the FLOW parameter is significantly correlated with a number of other effluents (phosphorous and nitrogen).

Figure 1. Pulp Output (Q), Waste Water Flow (FLOW), Biological Oxygen Demand (BOD) (mean values)



5.1 The First-Stage Results

The nonparametric linear programming problem *(4.2)* was computed by using all 152 observations at one time. The results are shown in *Table 1.* When the distance measures were computed by taking each year separately (i.e. 19 sets with 8 observations each) almost all the distance measures were equal to 1, due to the small number of observations. On the other hand, when the *sequential approach* (see Lovell 1993, pp. 47-49), which allows progressive technical change, was used to compute the distance measures, the results did not differ greatly from those presented in *Table 1.* However, we chose to use the pooled data in the nonparametric analysis, and to allow the period- and plant- specific effects to enter the stochastic model. This has the advantage that one can test explicitly whether they should be included in the model.

Table 1. Distance Measures from the Nonparametric Linear Programming Model

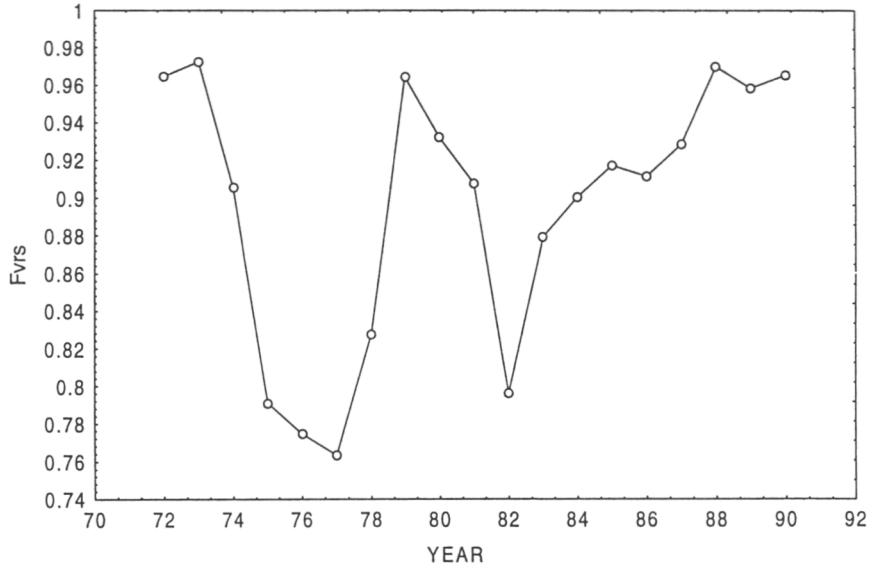
VARIABLE	MEAN	ST. DEV.	MIN	MAX	SKEW.	KURT.
F^{vrs}	0.90	0.12	0.54	1	-0.99	2.85
F^{crs}	0.87	0.13	0.51	1	-0.87	2.76
F^{vrs} plant 1	0.91	0.14	0.58	1		
F^{vrs} plant 2	0.92	0.10	0.60	1		
F^{vrs} plant 3	0.75	0.12	0.60	1		
F^{vrs} plant 4	0.94	0.10	0.73	1		
F^{vrs} plant 5	0.86	0.12	0.54	1		
F^{vrs} plant 6	0.87	0.11	0.70	1		
F^{vrs} plant 7	0.94	0.08	0.76	1		
F^{vrs} plant 8	0.97	0.06	0.81	1		

The results for the mean value for the whole sample and the mean values for each plant over time are summarized in *Table.1.* Besides providing the measures for the distances from the frontier, the results are interesting in their own right, since they can be

interpreted as efficiency measures (see e.g. Färe et al. 1989). The distance scores were computed for both the constant-returns-to-scale (F^{crs}) and variable returns to scale (F^{vrs}) models. The differences in the distance scores between the two models are small, the Spearman rank order correlation coefficient between the two series is 0.91 and the correlation is significant at the 1 % level. The mean efficiency in the variable returns to scale is 0.90. Thus, the sulphate pulp production for plants in the sample could be increased by about 10 percent on average if all plants were to operate on the production frontier. In the constant-returns-to-scale case, the mean efficiency is 0.87, implying that the current output could be produced with 13% lower cost. These results indicate that there are considerable gains to be made by improving the efficiency.

The variation of the mean efficiency over time is shown in *Figure 2*. The changes in efficiency over time can be roughly divided into two periods. The average inefficiencies have

Figure 2. Distance measures (F^{vrs}) from the Nonparametric Linear Programming Model



been the largest in 1973-77 and 1980-82, which are periods that coincide with economic slumps. In these years the plants' outputs were below their long-run trends. Thus, it appears that the adjustment to changes in the market environment happens partly through changes

in efficiency. However, it should be stressed that the above efficiency scores here are mainly a means to an end, i.e. they are used as the dependent variable in the estimation of the stochastic distance function. From this standpoint, the results show that had we set the dependent variable equal to 1 in the stochastic output distance function model (i.e. used the frontier approach), we would have obtained a rather poor approximate of the actual distance measures (59 of the 152 observations have distance measure equal to 1).

The distance scores belong to the interval (0,1) with concentration at the upper bound, which causes the distribution to be negatively skewed (see *Table 1*). This makes ordinary least squares (OLS) regression analysis an inappropriate technique, since the results would be biased and inconsistent. However, due to the linear homogeneity of the output distance function of degree +1 in outputs, for each observation, the left-hand side and right-hand side output variables are multiplied by $\lambda = 1 / Y_4$, where Y_4 is the output of one of the bads (suspended solids). After imposing the homogeneity restriction on this transformation and employing the logarithmic transformation, the histogram of the dependent variable and the *Jarque-Bera-statistics* (= 1.82) indicated that the transformed variable is normally distributed (skewness -0.26 and kurtuosis 3.09). Thus, OLS can produce consistent and unbiased estimates. A similar transformation has been used by Lovell et al. (1990) and Grosskopf et al. (1996). However, in the present study the transformation is computed on actual distance scores, not on a constant value of one.

For the purpose of comparison, we also experimented with a model, in which only the logarithmic transformation was imposed. Thus, the dependent variable belongs to the interval ($-\infty$ and 0) with concentration at the upper bound. To resolve this problem, we assume that the dependent variable is censored at zero and apply a Tobit model and maximum likelihood estimation (see Greene 1993).

5.2 The Second-Stage Results

The estimation of the stochastic output distance function consisted of first specifying the preferred panel data model. In particular, five different panel data specifications were estimated: 1) the pooled model without plant- and period-effects (OLS), 2) a model with fixed

plant indicators (OLS), 3) a model with random plant indicators (generalized least squares GLS), 4) a model with plant- and period-specific fixed indicators (OLS), and 5) a model with plant- and period-specific random effects (GLS). Next, in order to analyze the sensitivity of the results to different specifications and estimation methods, the output distance function was estimated using 16 different specifications. These model specifications were motivated by the following issues. First, our interest was in comparing three different estimation approaches: the two-stage approach estimated using OLS, the "conventional approach" using OLS (Lovell et al. 1990, Grosskopf et al. 1996), and the Tobit maximum likelihood method. Secondly, two different functional forms were compared, namely, the Cobb-Douglas and restricted translog. Thirdly, due to possible multicollinearity and functional dependence between FLOW and BOD, experiments were done with models in which only one at a time of these bad outputs was included. Finally, since the estimation period is rather long (19 years), and as environmental regulations have become more strict, the plants have been able to adjust their production processes, one would expect that these changes would also have an impact on the shadow prices. Thus, the estimations were also carried out for two sub-periods, namely, 1972-80 and 1981-90. The estimations of equation *(4.3)* were computed using the *LIMDEP 7.0* program.

Turning to the results, shown in *Table 2* and *Table 3*, it should first be noted that they were not very sensitive to whether the DF^{vrs} or DF^{crs} was used as a dependent variable. This is not surprising in light of the fact that the simple Spearman rank order correlation between these two distance scores is 0.91. Consequently, and for simplicity, the estimations are presented only for the models in which the DF^{crs} distance scores were used as a dependent variable. The test statistics used to discriminate between the different panel data specifications (Log Likelihood Ratio (LR) and Hausman tests) indicated that the two-factor fixed effects (TFFE) model described the data generating mechanism best (*see* Appendix ESII.2). The fixed effects account for exogenous shocks and omitted variables. For example, the plant-specific effects may capture effects of inputs that are invariant over time but vary across plants (e.g. type of machinery), whereas the period-specific effects may reflect changes in the macroeconomic environment. Although the sulfate pulp plants form a relatively homogeneous group, they are known to use e.g. different types of capital and wood raw material, which are likely to be captured in the plant-specific fixed effects. Also, the

*Table 2. Estimated Parameters for Alternative Models (fixed effects dummy variables included; 152 obs.).**

Parameter	OLS Cobb-Douglas	Tobit Cobb-Douglas	OLS Rest. Translog	Tobit Rest. Translog	OLS Rest. Translog	OLS Rest. Translog
α_0	1.736 (2.5)	2.066 (2.7)	-1.148 (1.4)	10.905 (1.4)	2.105 (3.1)	1.434 (2.1)
γ_K	-0.103 (2.6)	-0.091 (2.0)	-0.119 (3.1)	-0.165 (5.9)	-0.125 (3.3)	-0.157 (4.3)
γ_L	-0.136 (2.9)	-0.112 (2.0)	-0.123 (2.7)	-0.300 (7.7)	-0.156 (3.5)	-0. 058 (1.6)
γ_M	-0.821 (14.8)	-0.688 (7.8)	-0.800 (14.8)	-0.660 (12.5)	-0.810 (15.0)	-0.829 (15.7)
α_Q	0.857 (17.0)	0.700 (6.9)	1.192 (6.1)	1.291 (1.9)	0.852 (17.8)	0.915 (29.7)
β_{FLOW}	0.135 (2.7)	0.095 (1.7)	-0.146 (1.0)	1.759 (1.9)	0.148 (3.1)	
β_{BOD}	-0.008 (0.4)	-0.010 (0.4)	-0.046 (0.5)	-1.768 (2.8)		0.085 (2.8)
α_{Q2}			-0.039 (1.7)	0.190 (3.5)	0.006 (2.7)	0.010 (2.9)
μ_{QFLOW}			-0.005 (0.4)	-0.574 (8.1)	-0.006 (2.7)	
μ_{QBOD}			0.045 (1.6)	0.073 (1.9)		-0.010 (2.9)
β_{FLOW2}			0.067 (2.2)	0.393 (8.8)	0.006 (2.7)	
$\beta_{FLOWBOD}$			-0.017 (2.0)	-0.326 (6.7)		
β_{BOD2}			0.017 (0.8)	0.258 (8.9)		0.010 (2.9)
$\bar{R}^2$	0.72		0.98		0.98	0.98
Log-L	166.87	89.64	175.50	160.65	171.38	167.52

**The White heteroskedasticity-consistent t-values are shown in parentheses. Note: The fixed effects dummy variables are omitted for simplicity.*

*Table 3. Estimated Parameters for Two Periods; 1972-80 and 1981-1990 (fixed effects dummy variables included; 152 obs.).**

Parameter	OLS Cobb-Douglas 1972-80	OLS Cobb-Douglas 1981-1990	OLS Rest. Translog 1972-80	OLS Rest. Translog 1972-80	OLS Rest. Translog 1981-90	Tobit Rest. Translog 1981-90
α_0	-0.077 (0.0)	-0.307 (0.2)	-0.095 (0.1)	-1.155 (0.5)	0.140 (0.1)	0.167 (0.1)
γ_K	-0.218 (3.3)	-0.205 (3.2)	-0.219 (3.5)	-0.189 (2.4)	-0.228 (3.6)	-0.229 (4.0)
γ_L	-0.094 (0.5)	-0.054 (1.1)	-0.068 (0.4)	0.167 (0.7)	-0.061 (1.3)	-0.059 (1.4)
γ_M	-0.599 (5.7)	-0.653 (8.5)	-0.610 (6.1)	-0.821 (7.1)	-0.656 (8.8)	-0.658 (8.7)
α_Q	0.517 (5.0)	0.995 (21.1)	0.511 (5.8)	0.784 (7.4)	0.991 (22.6)	0.984 (41.1)
β_{FLOW}	0.530 (5.6)	0.012 (0.3)	0.489 (5.5)		0.009 (0.2)	
β_{BOD}	-0.047 (0.7)	-0.007 (0.5)		0.216 (2.0)		0.016 (0.7)
α_{Q2}			0.007 (1.8)	0.018 (2.4)	0.003 (1.4)	0.004 (1.3)
μ_{QFLOW}			-0.007 (1.8)		-0.003 (1.4)	
μ_{QBOD}				-0.018 (2.4)		-0.004 (1.3)
β_{FLOW2}			0.007 (1.8)		0.003 (1.4)	
$\beta_{FLOWBOD}$						
β_{BOD2}				0.018 (2.4)		0.004 (1.3)
$\bar{R}^2$	0.99	0.99	0.99	0.98	0.99	0.99
Log-L	84.13	131.23	86.11	70.67	132.37	132.24

**The White heteroskedasticity-consistent t-values are shown in parentheses. Note: The fixed effects dummy variables are omitted for simplicity.*

period-specific effects capture the impact of e.g. the oil crises shocks, which should not be mixed up with efficiency effects.

The estimation results in *Table 2* and *Table 3* show that the parameters and their standard errors and thus the t-values are to some extent sensitive to the particular specification and estimation method used. However, some caveats should be made as to the Tobit model results. The starting values for the Tobit maximum likelihood estimations are based on the first stage OLS values. Due to the nature of the Tobit model and the program used, the homogeneity restriction that the output values sum to one, could be imposed only on the OLS values. Thus, for the final maximum likelihood results, this restriction is not satisfied precisely, and although the results are useful for comparative purposes, they do not strictly satisfy the theoretical restrictions.

The *t*-values show that for the Cobb-Douglas specification all the coefficients are significant at the 5% level, except the ones for BOD (note also that in the Tobit model, FLOW is significant at the 10 % significance level). The diagnostics and specification tests reported in Appendix ESII.1 provide further information on the robustness of the models. The results indicate that the restricted translog specification with only the FLOW bad output included (i.e. BOD omitted) should be chosen over the models where both the FLOW and BOD are include. According to the functional form test (LR-test), the Cobb-Douglas specification is rejected in favor of the restricted translog form. Also, the Hausman test indicated that the right hand-side output variables can be regarded as exogenous. However, rather than going into more detail on the different specification tests, we turn to an evaluation of how the different specifications effect the shadow price results, shown in *Table 4*.

Table 4 provides the mean shadow prices (i.e evaluated at mean of the data) for the 16 different model specifications. The absolute shadow prices for undesirable outputs were computed using equation *(3.4)* and the assumption that the absolute shadow price of pulp (r_Q) is equal to its observed market price (r_Q^o). It is important to note that the figures give the shadow prices at the actual emission levels, not at the level of some pollution regulation. This is indeed a desirable feature, since in practice the plants have rarely met the constraints exactly.[6] Usually the plants have been below the constraint, probably due to the fact that they start to adjust to a new regulation already when they receive prior information

about a forthcoming regulation, even though it may actually come into force in 1-5 years. Thus, the shadow prices do not measure the effect of the regulation but rather the marginal rate of transformation between the good and bad outputs.

Table 4. The Absolute Shadow Prices for Undesirable Outputs (FIM/ton)

MODEL	FLOW	BOD
1. Cobb-Douglas OLS two-stage	429.9 (2.4)	-25.5 (0.4)
2. Cobb-Douglas conventional	406.7 (2.3)	-49.1 (0.7)
3. Cobb-Douglas TOBIT	370.4	-38.9
4. Cobb-Douglas OLS 1972-80	2798.8	-248.1
5. Cobb-Douglas OLS 1981-90	32.9	-5.1
6. Rest. Translog OLS	-336.3	-34.5
7. Rest. Translog TOBIT	5504.0	2949.5
8. Rest. Translog conventional	318.8	-189.4
9. Rest. Translog OLS no BOD	413.5	
10. Rest. Translog OLS no BOD conventional	396.4	
11. Rest. Translog OLS no FLOW		124.9
12. Rest. Translog OLS no FLOW conventional		-30.3
13. Rest. Translog OLS no BOD 1972-80	2570.7	
14. Rest. Translog OLS no BOD 1972-80	-4.3	
15. Rest. Translog OLS no BOD 1972-80		500.4
16. Rest. Translog OLS no BOD 1972-80		-5.9

Note: The shadow prices are evaluated at the mean of the data. For the stochastic Cobb-Douglas models 1 and 2, the t-values of the shadow prices are shown in parentheses.

For the Cobb-Douglas OLS models (models 1 and 2) significance tests were done on the shadow prices (for the restricted translog specification this is difficult to compute, due to the second order terms). Since the tests are on a single non-linear hypothesis, the small sample distribution of the test statistic is unknown. However, the test statistic is asymptotically distributed as χ^2_q , where q is the number of hypotheses. According to the *t*-, *F*-, -and *Wald* tests, the null hypothesis that the shadow price estiamtes are zero could be rejected for the FLOW shadow price in both model 1 and model 2 at the 5% significance level. However, for the BOD shadow price the null hypothesis could not be rejected for either of the models. Therefore, the BOD shadow prices are used in the present case mainly to compare the relative performance of the different approaches, rather than to draw policy conclusions.

The most significant feature of the above shadow price estimates is the fact that they are positive for FLOW in most of the specifications. The exception is the negative shadow price of FLOW in models 6 and 12. The positive shadow prices indicate that the marginal reductions of FLOW have been associated with increases in plant revenues. On the other hand, in most specifications the shadow prices for BOD are negative, thus, indicating the trade-off between BOD abatement and revenues. Indeed, the results from other studies on shadow prices of pulp and paper mills, and the economic textbook would lead us to expect negative shadow prices for undesirable outputs (e.g. Pittman 1981 and Färe et al. 1993). On the other hand, already in their pioneering study of water pollution, Kneese and Bower (1968) noted that pulp and paper plants have incentives other than environmental concerns to reduce their effluents. In particular, internal process changes which increase the recirculation of waste water have been adopted for the purpose of increasing the efficiency of the production process. Kneese and Bower (1968) state that *"Changes in production processes, however, have the major impact on wastes generation, although it should be emphasised that technological changes have not been, and are not usually, instituted because of water quality problems. In fact, most changes in production technology have been stimulated by factors unrelated to water problems, and have been developed without explicit considerations for their effects on water quality" (pp. 44-45).*

There are basically two ways for the plants to reduce waste water: by modifying the effluents using external treatment measures (e.g. building an aerated pond or activated

sludge plant) and/or by reducing the generation of effluents by implementing internal production process changes. External treatment measures clearly add extra costs to the plants and do not generate any additional benefits in terms of higher efficiency or productivity. Although the plants have received investment credits from the state for pollution abatement equipment, these have covered only a small portion of the total abatement costs. Consequently, 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 Finnish pulp plants in developing the production process has been to aim at closed-loop water systems, which simultaneously improve the efficiency of the control of production systems and reduce water pollution. As a result of this strategy, the production of one ton of pulp in 1990 required on average ten times less water than in the 1950s.[7] On the other hand, the negative shadow prices for BOD are most likely the result of the fact that the BOD effluents have to a large extent been reduced by end-of-pipe measures, such as by building biological treatment lagoons.

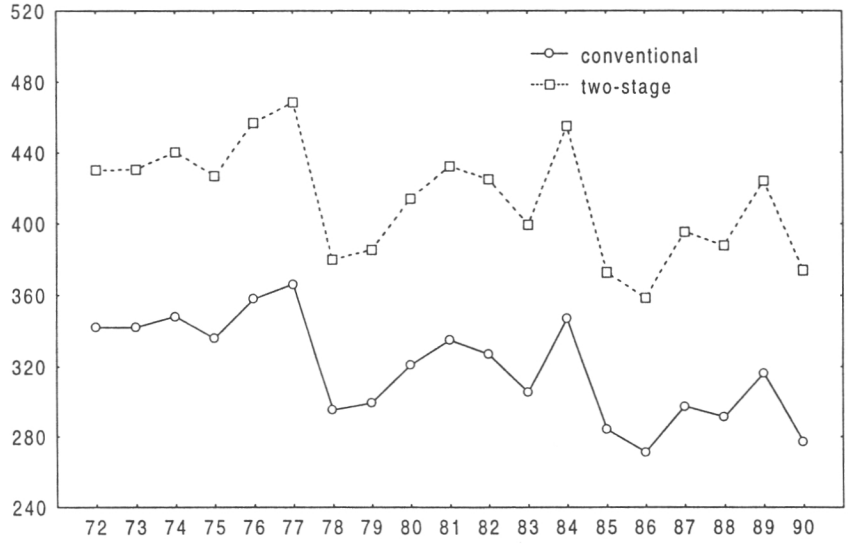
How do the shadow prices vary between the different model specifications and estimation approaches? First, *Table 4* shows that the shadow prices are sensitive to the specific functional form used in the estimation. The OLS and Tobit estimations produce similar results for the Cobb-Douglas specification, but the results are quite different for the restricted translog specifications. The relatively large absolute shadow price values generated by the Tobit model for restricted translog specification most probably reflect the fact that the output distance function property of being homogeneous of degree +1 in outputs has not been imposed on the Tobit model (due to computational difficulties). The impact of not imposing this restriction is naturally much larger for the translog specification than for Cobb-Douglas specification due to the second order terms.

The "conventional approach" and the two-stage approach produce rather similar results for the FLOW and BOD shadow prices for the Cobb-Douglas specifications (models 1 and 2). Also, for the restricted translog models, in which the BOD bad output is omitted (models 9 and 10), the FLOW shadow prices are of the same magnitude. The greatest differences between the two approaches are produced by the restricted translog specifications, where both FLOW and BOD are included. Indeed, the two-stage approac

generates for this specification a FLOW shadow price that is of equal size than that generated by the conventional "single-stage" approach, but with opposite sign. The results for the same functional specification in which the BOD variable is omitted indicates that the cause for this discrepancy is the correlation between FLOW and BOD variables. Why the two-stage approach is more effected by this multicollinearity than the conventional approach is unclear.

Figure 3 shows the absolute shadow prices for FLOW, computed over time using the two-stage approach (model 9) and the conventional approach (model 10). The similar pattern of the shadow price reflects to a large extent the fact that, due to the method of computation (i.e. equation *(3.4)*) , the shadow prices vary in line with variations in the data. The FLOW shadow prices have a downward trend, the two-stage approach generating somewhat higher shadow prices relative to the conventional approach.

Figure 3. Absolute Shadow Prices of FLOW (FIM/ton), evaluated at the mean values of the data, 1972-90.



Finally, from the sub-period results, two features stand out clearly. First, the estimates from the sub-period 1972-80 tend to be significantly higher in absolute terms than for the whole period or the later sub-period. Secondly, and in contrast, the estimates from the sub-

period 1981-90 are very much lower than the estimates for the whole period or the earlier sub-period. Whether this result reflects the possibility that the abatement measures and internal process changes have been relatively smaller at the end of the study period, and consequently their impacts on production and revenues have been likewise smaller, cannot be judged on the basis of the present study.

The results of the present study question the appropriateness of imposing the shadow prices to be non-positive a prior, as e.g. Färe et al. (1993) and Coggins and Swinton (1996) have done. Neither of these two studies provide arguments of why the shadow prices of undesirable outputs must be non-positive. The axioms on which the output distance function is based do not require such a restriction. If the shadow prices could be interpreted strictly to reflect the impact of environmental regulation, the restriction could be defended. However, as Färe et al. (1993) note *"The approach employed here does not require information on regulatory constraints; shadow prices reflect the trade-off between desirable and undesirable outputs at the actual mix of outputs which may or may not be consistent with the maximum allowable under regulation"* (p. 376). As argued above, there are good reasons, at least in the pulp and paper industry (which is also the industry studied by Färe et al. 1993), to expect this trade-off to be positive.

6. Conclusions and Implications

The distinguishing feature of the present study is the combination of nonparametric linear programming with the stochastic distance function and with plant level panel data. This framework provides an alternative method to previous deterministic and stochastic distance function studies. First, the estimation procedure does not require the plants to operate on the frontier of the production technology but is instead based on the actual distance measures. In the unlikely case that all the distance scores turn out to be equal to one, the present approach reduces to the conventional approach. Secondly, no a priori restrictions are applied to the values of the undesirable output parameters. Furthermore, plant level panel data from a relatively homogeneous industry sector provide more informative and robust results than the commonly used aggregate cross-section or time series data.

The results of the present study show that the reduction of FLOW in Finnish pulp plants has enhanced the revenues of the plants. This result should not, however, 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 that has caused these plants to reduce water pollution; another is the fact that pollution control measures and improvements in the production process appear to be strongly positively correlated.

By contrast, the abatement of BOD appears to have a negative opportunity cost (in terms of foregone revenues) to the plants. This result is probably due to the fact that the abatement of BOD has been achieved to a large extent by end-of-pipe measures, which add to production costs, and that these measures do not have any positive spill-over impacts on the production (productivity) of pulp, as do the internal process change measures (re-circulation of waste water).

How legitimately the above results can be generalized to other production processes is an empirical question. Nevertheless, the result indicates that one should not a priori rule out the possibility that pollution control may be positively correlated with an increase in firms' revenues. The literature on the impact of environmental regulations and pollution control on firms has often assumed that these measures will have only negative opportunity costs on firms production. However, the present study has demonstrated that pollution control may be a side effect of firms' interest in improving production efficiency.

Finally, there are a number of ways to improve the research in this area, of which some are mentioned here. First, the properties of the estimates computed using the two-stage approach should be compared to the estimates computed using the conventional "one-stage" approach. This comparison could be implemented e.g. by using Monte Carlo simulation analysis. Secondly, Tobit estimations should be extended to allow for theoretical restrictions to be imposed in the maximum likelihood estimation. Finally, it would be important to evaluate the curvature and monotonicity of the output distance function at each data point.

Notes

1. According to Shephard (1974, p. 205), *"... , for the future where unwanted outputs of technology are not likely to be freely disposable, it is inadvisable to enforce free disposal of inputs and outputs. Since the production function is a technological statement, all outputs, whether economic goods are wanted or not, should be spanned by the output vector y."*
2. There are reasons related, e.g. to positive spillover effects of pollution control which suggest that shadow prices of bads may also be positive. Gray and Shadbegian (1993) note that *"In some cases regulations may increase productivity. In response to pressures to reduce waste water discharges, some plants adopted "closed-loop" production processes and discovered after doing so that the cost savings from recycling raw materials reduced total costs. .....New equipment, installed to reduce pollution, may also be more productive than the old equipment it replaces."* Similar evidence can also be found in the classical study by Kneese and Bower (1968) (see also Oates et al. 1993).
3. 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 $v \leq y \in P(x)$ then $v \in P(x)$. Basically, weak disposability implies that radial (equiproportional) reduction in outputs is possible, but reduction of some outputs may not be feasible without altering inputs. On the other hand, strong disposability implies that outputs can be disposed without any resource use. For a more detailed description of the concepts, see Färe et al. (1994).
4. Although an inefficient observation has no supporting hyperplane (since there is no supporting frontier), shadow prices for inefficient observation still make sense in the present context. The way to look at this is that, although the observation is inefficient, we calculate the shadow price that would have obtained, if the observation had been efficient. The inefficient observation is scaled *proportionally* up to the frontier. Then the derivatives, which give the shadow prices, yield the same mutual relation as the derivatives evaluated at the optimal (efficient) point. Thus, as long as the scaling (inefficiency) is proportional, it does not effect the relations between the shadow prices.
5. The program used for computing the distance scores is based on Fortran, and it utilizes IMSL math/library's DDLPRS-subroutine (see IMSL, pp. 888-891).
6. Pittman (1981, p. 9) states that *"given the imprecise nature of production and pollution control technologies, it seems unlikely that one would ever find such a constraint exactly met. This would not seem to imply, however, that such plants are not behaving under constraint, nor that they would attach zero value to a relaxation of the constraint."* Similarly, Brännlund and Löfgren (1996) note that in the case of the Swedish pulp industry, *"Most plants have an average emission level far below the allowed level. We do not think, however, that this means that almost all kinds of regulations are ineffective. Instead we believe that the main reason for this inequality is that the firm cannot control its emissions exactly. In other words, the plant's waste load is subject to stochastic fluctuations."*
7. The fact that the internal process measures which reduce waste water also have positive effects on the revenues of the plants is present also in the data collected from plants by the water authority. The data concerning the net costs of the internal process changes undertaken due to pollution regulation show that for some plants and some years the net effects are positive.

References

Banker, R.D. and W.W. Cooper (1994). Validation and Generalization of DEA and its Uses, to be published in *Trabajos de Investigacion Operativa*.

Barbera, A.J. and McConnell, V.D. (1990). The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects, *Journal of Environmental Economics and Management* 18, 50-65.

Brännlund, R. (1996). Estimating Shadow Prices of Undesirables, unpublished mimeo, Department of Economics, University of Umeå

Brännlund, R. and K.-G., Löfgren (1996). Emission Standards and Stochastic Waste Load, *Land Economics*; 218-30..

Coggins, J.S. and J. R. Swinton (1996). The Price of Pollution: A Dual Approach to Valuing SO2 Allowances, *Journal of Environmental Economics and Management* 30: 58-72..

Conrad, K. and Morrison, C.J. (1989): The Impact of Pollution Abatement Investment on Productivity Change: An Empirical Comparison of the U.S., Germany, and Canada. *Southern Economic Journal*, January: 684-698.

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

Färe, R. (1988): *Fundamentals of Production Theory*, Lecture Notes in Economics and Mathematical Systems, vol. 311, Berlin: Springer Verlag.

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

Gollop, F.M. and M.I. Roberts (1983). Environmental Regulations and Productivity Growth: The Case of Fossil-fueled Electric Generation, *Journal of Political Economy*: 654-674.

Gray, W.B. & S.J. Shadbegian (1993). Environmental Regulation and Manufacturing Productivity at the Plant Level, National Bureau of Economic Research, Working Paper No. 4321.

Greene, W. (1993). *Econometric Analysis*, 2nd ed., MacMillan.

Grosskopf, S. (1996). Statistical Inference and Nonparametric Efficiency: A Selective Survey, *in Olesen, Petersen and Lovell (eds. 1996).*

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

Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1996). Budget Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling, *Review of Economics and Statistics*, forthcoming.

Hetemäki, L. (1994). The Impact of Pollution Control on a Firm's Production Technology and Efficiency: A Stochastic Distance Function Approach, Proceedings of Environmental Economics Conference held at Ulvön (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. (1995). The Impact of Pollution Control on Technology and Efficiency: A Stochastic Distance Function Approach, in Boero, G. and Silberston, Z.A. (eds.), *Environmantal Economics*, Macmillan.

Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge University Press.

IMSL MATH/LIBRARY. Fortran Subroutines for Mathematical Applications. Version 1. IMSL, Inc. 1987.

Jorgenson, D. and P.J. Wilcoxen (1990). Environmental Regulation and US. Economic Growth, *RAND Journal of Economics*, vol.21,no.2; 314-340.

Kneese, A.V. and B.T. Bower (1968). *Managing Water Quality: Economics, Technology, Institutions*. Resources for the Future, John Hopkins Press.

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.)

Myers, J.G. and L. Nakamura (1980). Energy and Pollution Effects on Productivity: A Putty-Clay Approach, in *New Developments in Productivity Measurement and Analysis, Kendrick* (ed.), University of Chicago Press.

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

Pittman, R.W. (1981). Issues in Pollution Interplant Cost Differences and Economies of Scale, *Land Economics*, 1-17.

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

Shephard, R.W. (1974). Comment on Diewert, E. "Applications of Duality Theory", in Intriligator, M.D. and Kendrick, D.A. (eds.), *Frontiers of Quantitative Economics vol. II*, North-Holland.

Simar, L. (1992). Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping, *Journal of Productivity Analysis*, no. 3; 171-203.

Appendix ESII. 1. Diagnostics and Specification Tests *

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) is accepted for the Cobb-Douglas model for 6 plants and for the restricted translog specification for 5 plants out of the total of 8 plants.

Heteroskedasticity. Whites's heteroskedasticity test, based on F-statistic was used. The computed F-statistic for the Cobb-Douglas specification is F = 2.76, and for restricted translog F = 1.46. Thus, homoskedasticty is rejected for the Cobb-Douglas at the 5% level, whereas it cannot be rejected for restricted translog at the 5% significance level. As a result, the White heteroskedasticity-consistent standard errors are used in the analysis for the parameters.

Normality. The *Jarque-Bera (JB) -statistic* was used to test whether the residuals are normally distributed. The null hypothesis of normality was rejected at the 5% significance level, for the Cobb-Douglas specification (*JB = 11.82*) and for restricted translog specification (*JB = 14.87*). However, when 4 outlier observations (obs. 46, 81, 82,120) were removed, the null hypothesis of normality is accepted (*JB = 3.35* and *JB = 3.37* for the Cobb-Douglas and restricted translog, respectively). It may be noted that the results for the shadow prices were not sensitive to estimating the models without the above four observations.

Orthogonality. The *Hausman specification statistic (H)* was used to test the exogeneity of the output terms on the right-hand side of the regression equation. The test indicated that (at the 5 % significance level) OLS is an efficient estimation method, in contrast to the instrumental variables estimator (H_3 = 2.67). The instruments used in the instrumental variables estimation were all the exogenous variables and in addition the following plant

level variables: the amount of electricity used, wages, nitrogen effluents, phosphorous effluents, gross value of output, and as an industry level variable the production price index.

Testing parameter restrictions. The *Wald test* of coefficient restrictions was carried out to see which of the functional forms is preferred and to examine whether the undesirable outputs (FLOW and BOD) should be included in the models. First, in the Cobb-Douglas specification, the F-statistic for testing that the coefficient of the BOD variable is zero, produced an F = 1.62. Thus, we would accept the null hypothesis at the 5 % significance level. Similar test of the coefficient of the FLOW variable produced an F = 8.91, thus we would reject the null at the 5% significance level. A test of Cobb-Douglas vs. restricted translog form produced an F = 4.63, thus indicating that one should reject the Cobb-Douglas in favor of the restricted translog at the 5 % significance level.

Functional form specification. The Ramsey RESET statistic was used to test the null of correct functional form specification of the original model against the alternative that the squared fitted values have been omitted. In the Cobb-Douglas and restricted translog specifications, the coefficient for the squared fitted term had t-values of 1.07 and 1.95, respectively, thus giving support to the original functional forms.

**See Appendix IV Test Statistic for a detailed description of the tests*

Appendix ESII. 2. Test Statistics for Panel Data Model Selection

COBB-DOUGLAS

Model	LR-test	Chi-sq.	d.f.	Decision
1. constant term only	2 vs 1	59.24	7	choose 2
2. plant and period fixed effects only	3 vs 1	40.93	6	choose 3
3. X -variables only	4 vs 1	79.67	12	choose 4
4. X and plant fixed effects	4 vs 2	20.43	5	choose 4
5. X and plant and period fixed effects	4 vs 3	38.74	7	choose 4
6. X and plant and period random effects	5 vs 4	115.6	18	choose 5
	5 vs 3	154.33	26	choose 5
	5 vs 6	7.77	6	choose 5

RESTRICTED TRANSLOG

Model	LR-test	Chi-sq.	d.f.	Decision
1. constant term only	2 vs 1	59.24	7	choose 2
2. plant and period fixed effects only	3 vs 1	69.66	12	choose 3
3. X -variables only	4 vs 1	117.44	18	choose 4
4. X and plant fixed effects	4 vs 2	58.20	11	choose 4
5. X and plant and period fixed effects	4 vs 3	47.78	7	choose 4
6. X and plant and period random effects	5 vs 4	133.50	18	choose 5
	5 vs 3	181.29	26	choose 5
	5 vs 6	7.73	10	choose 5

ESSAY III

Do Environmental Regulations Increase Production Efficiency? Evidence From The Pulp Industry

Lauri Hetemäki

Abstract

According to the so-called *Porter hypothesis,* properly designed environmental standards can trigger innovation and production efficiency gains that may lead to absolute advantages over non-regulated firms. This hypothesis has recently been the subject of active debate in the environmental economics literature. In order to provide a framework of empirical analysis for testing such a hypothesis, this paper examines the impact of environmental regulations and pollution abatement measures on the production efficiency of Finnish pulp producers. An output distance function model and the stochastic frontier estimation approach, based on microlevel panel data, are used to estimate the technical efficiency of pulp plants. The model simultaneously estimates technical inefficiency and the determinants of such inefficiency. The results indicate that environmental regulation has reduced the production efficiency of pulp plants.

I am grateful to helpful comments offered by Scott Atkinson, Tim Coelli, Mika Hakuni and participants of the 3rd Ulvön Conference on Environmental Economics, June 18-20, 1996. I would also like to thank Tim Coelli for providing his *FRONTIER 4.1* program. This research was supported by the Academy of Finland and the Nordic Council of Ministers Project "Economic Instruments in Environmental Policy: The Nordic Forest Sector".

1. Introduction

In recent years, an active debate has emerged on the impact of environmental regulations and pollution abatement on the production efficiency and competitiveness of the firm. The issue has aroused considerable interest among economists and policymakers. The discussion started with Porter's studies (1990, 1991), in which he formulated his hypothesis that environmental regulations can improve firms' overall production efficiency, and thus competitiveness, relative to firms not regulated. More recently, Porter and van der Linde (1995, p. 98) argue, *"that properly designed environmental standards can trigger innovation that may partially or more than fully offset the costs of complying with them. Such "innovation offsets", as we call them, can not only lower the net costs of meeting environmental regulations, but even lead to absolute advantages."* The authors further contend that innovation offsets occur mainly because pollution reduction is often coincident with improved efficiency of resource usage. Thus, the hypothesis rests heavily on the inference that stiffer environmental regulations result in greater production efficiency.

Although Porter's hypothesis has been received enthusiastically, especially in management literature and among politicians, economists have tended to be critical of its theoretical and empirical foundations (e.g. Jaffe et al. 1995, Palmer et al. 1995, Jaffe and Palmer 1996, Simpson and Bradford 1996). Indeed, the economic textbook version states that introducing environmental regulation to areas it has not previously been applied or tightening existing regulations inevitably results in lower profits for the firm. However, Simpson and Bradford (1996) provide a theoretical (strategic trade) model in which it is possible that an effluent tax can lead to a competitive advantage for the regulated firm compared to non-regulated firms. But as the authors state, this result is rather an exception than a general implication of their model. Palmer et al. (1995) further observe that indirect empirical evidence provides no support for Porter's hypothesis. However, as Jaffe and Palmer (1996) indicate, apparently no study has been done that directly tests the hypothesis. Indeed, according to Jaffe and Palmer (1996), there is still some ambiguity as to exactly what the Porter hypothesis is. The state of the discussion has recently been summarized in the *Journal of Economic Perspectives* (Porter and van der Linde 1995, Palmer, Oates and Portney 1995).

Although Palmer et al. (1995) clearly do not accept the basic arguments of Porter's hypothesis, they do agree that environmental regulations and production

efficiency may be related. According to Palmer et al. (p. 120), *"we acknowledge that regulations have sometimes led to the discovery of cost-saving or quality-improving innovation; in other words, we do not believe that firms are ever-vigilantly perched on their efficiency frontiers"*. However, Palmer et al. indicate that more systematic studies are needed to establish the extent of this effect.

Indeed, the empirical literature on the relationship between environmental regulations and production efficiency is still rather scarce. Although there have been a number of studies dealing with environmental regulations and productivity, few analyse the efficiency aspect, especially at the firm or plant level. Studies which have focused on efficiency and made use of microdata have been based on a *nonparametric linear programming approach* (e.g. Färe; Grosskopf, Lovell and Pasurka 1989, Bernstein et al. 1990, Yaisawarng & Klein 1994). Further, the extensive body of literature on *stochastic frontier functions* apparently does not include a single study that explicitly takes into accounts pollution or environmental regulations in the estimation of production efficiency (for a recent survey of stochastic frontier literature, see Greene 1993). This may be a significant shortcoming, since a large part of the stochastic frontier literature is concerned with production processes generating substantial amounts of waste and which are regulated (e.g. power generation, pulp and paper production, base metals and steel production, etc.). It is implicitly assumed that undesirable outputs and/or environmental regulations do not have an impact on the measurement of efficiency. Nonetheless, the results from the nonparametric linear programming studies which have explicitly incorporated undesirable outputs and environmental regulations in the production technology model indicate that the neglect of these factors can greatly bias the production efficiency measures (see the studies cited above).

Finally, we extend the existing literature on the impact of water pollution abatement and environmental regulations on pulp industry production. Previous studies have examined shadow prices for regulated undesirable outputs of pulp mills (e.g. Pittman 1981, 1983, Färe et al. 1993) as well as the impact of regulation on profitability (Brännlund et al. 1995) and productivity (Färe et al. 1989) of pulp mills. The latter studies do not include a regulation variable directly in the analysis but use instead the indirect approach discussed above. The present study extends this literature by examining the impact of the environmental regulation intensity variable on the production efficiency of pulp mills and by relating the findings to the Porter hypothesis. Depending on whether this impact is negative or positive, it can be interpreted as evidence that either rejects or

supports a central premise of Porter's hypothesis vis-à-vis the pulp industry. The present study also examines the shadow prices of both non-regulated undesirable outputs and regulated undesirable outputs. As will be shown below, the distinction between these two categories of pollution outputs turns out to be important in the case of the pulp and paper industry.

1.1 The Approach

In examining the impact of environmental regulation and pollution abatement on the production efficiency of the Finnish pulp industry, this study uses the output distance function as an analytical tool. The theoretical framework employed was first applied to environmental control problems by Färe et al. (1993) and later e.g. by Hetemäki (1995) and Coggins and Swinton (1996). The output distance function, like the more conventional production, cost and profit functions, provides a complete functional representation of the production technology. For the present purpose, it has some useful which the latter functions lack. First, the output distance function readily models multiple output (good and bad) production technologies without requiring price data. This is especially useful regarding production processes with bad outputs that do not have market prices. Moreover, it has been shown that the reciprocal of the output distance function is the output-based *Farrell measure of technical efficiency*. The output distance function has the additional benefit in that it is dual to the revenue function. This in turn implies that one can retrieve the shadow prices of outputs directly from the distance function by applying a dual Shephard's lemma.

The stochastic output distance function used here is estimated simultaneously with a model that explains why the plants may fail to reach the production frontier. In other words, what causes the plants to be inefficient? The determinants of inefficiency are estimated using the framework recently introduced by Battese and Coelli (1995) in the production function context. This approach allows us to analyse explicitly the impact of environmental regulations and pollution abatement on production efficiency.

The present study is based on plant-level balanced panel data, covering eight Finnish sulfate pulp mills observed annually for 19 years (1972-90). The sulfate pulp plants form a homogeneous group, for which the existence of a common production technology is a realistic approximation. It should be stressed that plant-level data is

essential in an environmental regulation impact analysis if regulations are set at the plant level (as is often the case).

The results of the study indicate that, contrary to Porter's hypothesis, environmental regulations have reduced the production efficiency of Finnish pulp producers. Also, the results for one of the regulated bad outputs shows unambiguously that the opportunity cost of reducing the bad is lower revenues. On the other hand, the shadow price of non-regulated bad output (waste water flow) was found to be positive, indicating that closed-loop water systems result in higher revenues.

The paper is organized as follows. Section 2 sets out the theoretical model; section 3 presents the econometric model; section 4 describes the institutional setting and data; section 5 derives and explains the results; and section 6 contains concluding remarks.

2. Theoretical Model

A detailed exposition of the theoretical background and various applications of distance functions is presented, e.g. in the recent textbook by Färe and Primont (1995). Empirical studies, which use the parametric output distance function to derive shadow prices for undesirable outputs (pollution) include Färe et al. (1993), Hetemäki (1995), and Coggins and Swinton (1996).

A reference technology, output set *P(x)*, is developed in this section for pulp plants which transforms factors of production $x = (x_1, x_{2,\ldots,} x_n) \in R_+^n$ into outputs $y = (y_1, y_2, \ldots, y_m) \in R_+^m$. The output set contains all technically feasible output vectors for the input vector x, i.e. $P(x) = \{y \in R_+^m : x \text{ can produce } y\}$. It is assumed that the technology satisfies the axioms (P.1-P.4) of Färe and Primont (1995, p.13). 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 (proportional) expansion of the output vector for a given input vector which is consistent with that output vector belonging to *P(x)*. The

assumptions for output set *P(x)* imply a set of properties for the output distance function (see Färe and Primont 1995, Chp. 2). In particular, the output distance function, if well-defined, will always be homogeneous of degree one in outputs. Given the properties of the output distance function, the following pair of relationships holds between the efficient frontier and the distance function:

$$(2.2a)\quad D_O(x,y) \le 1 \Leftrightarrow y \in P(x)$$

$$(2.2b)\quad D_O(x,y) = 1 \Leftrightarrow y \in IsoqP(x),$$

where *IsoqP(x)* is the boundary (frontier) of the output set. Thus the value of the output distance function must be less than or equal to one ($D_O \le 1$) for feasible output. The axioms regarding the output set *P(x)* impose a set of properties on the output distance function (for a detailed description, see Färe and Primont 1995). 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,

$$(2.3a)\quad R(x,r) = \max_{y}\{ry: D_O(x,y) \le 1\}$$

$$(2.3b)\quad D_O(x,y) = \max_{r}\{ry: R(x,r) \le 1\},$$

where $r \in R_+^M$ is a vector of output prices, and *R(x,r)* is the revenue function. Following the analysis of Färe et al. (1993), it can be shown that the relative output shadow prices can be derived as the derivative of the distance function using a dual Shephard's lemma. The ratio of the relative output shadow prices will reflect the slope of the production possibility frontier at the observed output mix. That is, for any two outputs, the slope of the production possibilities frontier (or the marginal rate of transformation, MRT) should equal the ratio of the corresponding output shadow prices or the marginal rate of transformation (for details, see Färe et al. 1993). Thus

$$(2.4)\quad MRT_{mm'} = \frac{\partial D_O(x,y)/\partial y_m}{\partial D_O(x,y)/\partial y_{m'}}$$

Assuming m denotes bad output and m' good output in equation (2.4), following Färe et al. (1993) and Coggins and Swinton (1996), $MRT_{mm'}$ can be interpreted to provide information on the marginal cost to plants of reducing their bad outputs, i.e. the cost of abatement. If the shadow price is negative, it indicates that the bad output (y_m) can be reduced only with an accompanying reduction of good output $(y_{m'})$. On the other hand, if the shadow price ratio is positive, the reduction of bad output is associated with an increase in good output (revenue).

3. Econometric Model and Determinants of Inefficiency

3.1 Stochastic Output Distance Function

There are basically three approaches to computing the output distance function: 1) nonparametric linear programming, 2) parametric linear programming, and 3) a stochastic, econometric model. In the present study the econometric approach is used, since it allows one to compute simultaneously the technical efficiency scores and the model for the determinants of inefficiency. Moreover, the purpose of the present study is to extend the existing stochastic frontier estimation literature by including explicitly the impacts of undesirable outputs on technical efficiency.

The stochastic formulation of the output distance function *(2.1)* can be expressed as

$$(3.1) \quad D_O = f(X, Y; \psi) \exp \varepsilon,$$

where D_O is the distance measure, $f(.)$ is the production technology, X is a matrix of inputs, Y is a matrix of outputs, Ψ is a vector of parameters to be estimated and ε is the random disturbance term intended to capture the effects of measurement error and statistical noise and is assumed to be independently and identically distributed as $N(0, \sigma_\varepsilon^2)$. The basic problem with the distance function, as concerns econometric estimation, is the inability to observe (obtain 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 logarithmically, the left-hand side of the distance function will be zero for all observations (i.e. $D_o = ln(1) = 0$).

To avoid the above problems, one can utilize the property that the output distance function is homogeneous of degree +1 in outputs (see Lovell et al. 1990 and Grosskopf et al. 1996). Thus the following relationship holds for the output distance function

$$(3.2) \quad \lambda D_o(X, Y) = D_o(X, \lambda Y).$$

Now let $\lambda = 1 / Y_m$ and write (*3.2*) as (*3.3*),

$$(3.3) \quad \frac{1}{Y_m} D_o(X, Y) = D_o(X, \frac{Y}{Y_m}).$$

From *(2.2)* we know that if $Y \in P(x)$, then $D_o(X,Y) \leq 1$. Consequently,

$$(3.4) \quad \frac{1}{Y_m} \geq D_o(X, \frac{Y}{Y_m}).$$

Equation *(3.4)* can be converted into a stochastic frontier model by parameterizing D_o and introducing the composed error structure, i.e. writing *(3.4)* as

$$(3.5) \quad ln\frac{1}{Y_{mpt}} = lnD_{opt}(X_{pt}, \frac{Y_{pt}}{Y_{mpt}}) + v_{pt} + u_{pt},$$

where $p = 1,....,P$ denotes the plants, and $t = 1,....,T$ denotes the time period, v represents random shocks and noise, and u represents the extent to which the plants fall below the output distance (production) frontier, i.e. it is a measure of production inefficiency. It is assumed that v_{pt} is independently and identically distributed as $N(0,\sigma_v^2)$, and u is assumed to be distributed independently of v and to satisfy $u_{pt} \leq 0$. After having estimated *(3.5)*, $E(u_{pt} | v_{pt} + u_{pt})$ is computed for each plant and period, from which plant-specific efficiency measures are calculated as

$$(3.6) \quad D_{opt}(X,Y) = exp\left\{-E(u_{pt} \mid v_{pt} + u_{pt})\right\}.$$

The composed error approach to estimate a stochastic distance function has previously been used by Grosskopf and Hayes (1993).[1] The composed error structure framework was originally formulated in a production function setting by Aigner et al (1977). In the present study, the above framework is extended by incorporating a model for u_{pt}, which allows one to estimate simultaneously the technical inefficiency (u_{pt}) and determinants of inefficiency.

3.2 A Model for Determinants of Technical Inefficiency

In order to estimate simultaneously the magnitude of inefficiency and the determinants of inefficiency, the framework proposed by Battese and Coelli (1995) in a production function setting is applied to the distance function framework. Let the u_{pt} in equation *(3.6)* be defined by equation *(3.7)*:

$$(3.7) \quad exp(-u_{pt}) = exp(-z_{pt}\delta - \omega_{pt}),$$

where the u_{pt}'s are assumed to be independently distributed, such that u_{pt} is obtained by truncation (at zero) of the normal distribution with mean $z_{pt}\delta$ and variance, σ^2; z_{pt} is a (*1 x h*) vector of plant-specific variables, which may vary over time; δ is an (*h x 1*) vector of unknown coefficients of the plant-specific inefficiency variables; and ω_{pt}, accounts for the unexplained or residual efficiency and is defined by truncation of the normal distribution with zero mean and variance σ^2, such that the point of truncation is $-z_{pt}\delta$, i.e. $\omega_{pt} \geq -z_{pt}\delta$. In other words, the z_{pt} variables shift the mean of the technical inefficiency error term.

In the above model, the measurement errors or technical inefficiency effects (u_{pt}'s) are assumed to be a function of a set of explanatory variables, the z_{pt}'s, and an unknown vector of coefficients, δ. These are factors that may not enter into the distance

function directly but that affect technical efficiency. The appropriate content and form of the Z vector is not obvious. The Z vector should reflect the reasons why inefficiency may arise, i.e. why the plants are not operating on the output distance function frontier. Here we examine two factors that may contribute to inefficiency. The factors reflect the intensity of the environmental regulation and capacity utilization rate (see below).

Following the approach of Battese and Coelli (1995), the parameters of equations *(3.5)* and *(3.7)* may be estimated simultaneously by the maximum likelihood method. The likelihood function is expressed in terms of the variance parameters, $\sigma_S^2 \equiv \sigma_V^2 + \sigma^2$ and $\gamma \equiv \sigma^2 / \sigma_S^2$. For the details and the derivation of the likelihood function and its partial derivatives with respect to the parameters of the model, see Battese and Coelli (1993).

4. The Institutional Setting

The Finnish pulp and paper industry has played a central role in the Finnish economy as well as being one of the country's primary sources of industrial pollution. The per capita export revenues earned by the Finnish pulp and paper industry are among the largest in the world (FIM 6,191 in 1990). However, it has also been the single major water polluting sector, and thus environmentalists and industrialists have often found a battle ground in the pulp and paper industry.

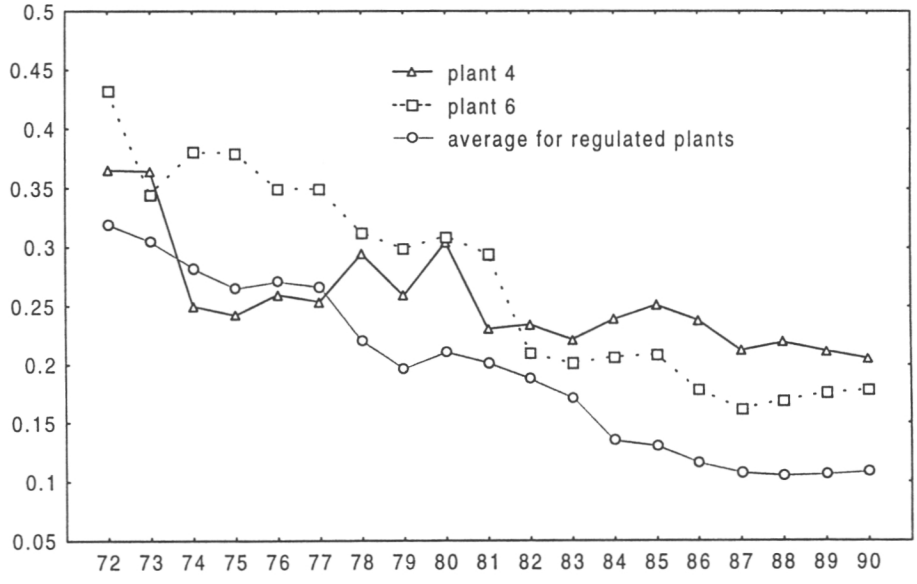
The present study examines only one sector of the pulp and paper industry, namely, sulfate pulp plants. The motive for this is that when the plants are operating in the same environment and producing the same product, one would hope to be able to obtain relatively satisfactory estimation results for the production technology. Moreover, sulfate pulp plants account for the bulk of waste water effluents generated by the pulp and paper industry. Finally, the sulfate pulp industry represents a typical process industry, whose inputs and end products, unlike many industries, are quite homogeneous. Thus the inputs and outputs can also be measured fairly precisely. The eight plants in the present sample accounted for about one-half of the total production of Finland's sulfate pulp industry during the period 1972-90.

The water pollution parameters subject to regulation are biological oxygen demand and suspended solids. The local water authority specifies effluent limits (standards) for BOD and SS as maximum allowed discharge, calculated as kg per ton of

produced pulp, or as tons/day (average per year or one or three months). Every plant has its own specific regulation on the substances that it discharges. Permits are typically renewed every 3-10 years. In practice, the application of regulations differs considerably among the plants. Also, two plants were never subject to regulation at any time during the study period (1972-90), and the remaining six plants were not all subject to regulation at some time in the 1970s. As a result of the above facts, the data on regulation varies considerably over time and across plants, thus allowing us to identify the impact of regulation on production.

In the case of the pulp and paper industry, it is also of particular interest to analyze the impact of the abatement of the total amount of waste water flow (FLOW) on production efficiency. Although, the water authority has *not regulated* FLOW, it has been reduced significantly over time. There are basically two reasons for this. First, FLOW, BOD and SS are to some extent linked to each other. Secondly, the reduction of FLOW is known to have positive spill-over effects on production efficiency. Below, we elaborate these points in more detail. In *Figure 1*, FLOW per ton of pulp produced in the pulp plants included in the present study is shown for the period 1972-90.

Figure 1. Waste Water Flow (FLOW) / Pulp Output (1000 m^3 /ton), 1972-90



Plants 4 and 6 are plants which have not been regulated with regard to on any water pollution substances, whereas the line denoted by "average for regulated plants" is the simple average for those plants which have been subject to regulations on BOD and suspended solids (SS). It may be observed that for both the regulated and non-regulated plants, waste water flow has decreased significantly during the period 1972-90.

Reducing water pollution from pulp plants can be accomplished either by reducing the generation of waste water (*pollution-prevention measures*) or by modifying the residual wastes (*end-of-pipe measures*). FLOW has been primarily reduced by pollution-prevention methods, such as in-plant recirculation of waste water. Since BOD and SS are "subsets" of FLOW, these substances have also been to some degree reduced by recirculation. However, the end-of-pipe measures or building of aerated treatment lagoons has been the major means by which BOD and SS have been reduced.

The pulp plants also have incentives to reduce the level of FLOW irrespective of environmental regulation, particularly because of production efficiency gains associated with closing the water system by recirculation. Indeed, a totally "closed-loop" water system has been a long-term objective in the pulp industry. This link between production efficiency and reduction of water pollution in the pulp and paper industry was already been noted in a pioneering study of water pollution by Kneese and Bower (1968).[2] Thus decreases in FLOW may result from any of the following three objectives: 1) increased production efficiency, 2) compliance with environmental regulations, 3) "green marketing strategies", or from a combination of these. It may be noted that the approach used in the present study does not allow for the analysis of the impact of green markets on effluents. Indeed, during the study period (1972-90), green marketing strategies were still of only minor importance for pulp plant operations; they have become more prevalent only recently.

It should be added that analysing FLOW separately from BOD and SS is also interesting in relation to the Porter hypothesis. In particular, Porter and van der Linde (1995) make the distinction between two classes of innovation offsets generated by environmental regulation. The first occurs when firms learn to deal with end-of-pipe pollution and the second is related to pollution-prevention measures. According to Porter and van der Linde (1995), it is innovations of the latter type that are central to the argument that environmental regulation can lead to production efficiency gains. Since there is no systematic data on which of the two types of measures pulp plants have adopted in response to new or stricter regulations, we cannot identify the impact of

regulation on abatement measures. However, as was stated above, FLOW has been primarily been reduced by means of pollution-prevention measures, whereas BOD and SS mainly by end-of-pipe measures. In summary, the above considerations lead us to expect that the shadow prices for FLOW may differ from those of BOD and SS.

A detailed description of the data is given in the Appendix II. The data consists of observations on quantity of sulfate pulp output (Q), net fixed capital stock (K), hours worked (L), value of materials input (M), and the level of waste water flow (FLOW), BOD and SS. The data used in the model of the determinants of inefficiency include capacity utilization rate (CU) of the pulp industry and a regulatory intensity variable (R). It may be noted that the regulatory intensity variable was also used by Gollop and Roberts (1983) to analyze the impact of environmental regulation on productivity. Further, it is precisely the intensity or stringency of the regulation, rather than regulation per se, that is key to Porter's argument. *"Stringent regulation can actually produce greater innovation and innovation offsets than lax regulation. Relatively lax regulation can be dealt with incrementally and without innovation, and often with "end-of-pipe" or secondary treatment solutions"* (Porter and van der Linde 1995, p. 100). In the present study, the regulatory intensity variable (R) is constructed from the data on plant-level standards imposed on BOD and SS (see Appendix ESIII.1 for details). R is given values between zero and one. It is unity when the water authority has not mandated any effluent standard for SS or BOD. The closer R is to zero, the more restrictive the regulation.

5. Estimation Procedure and Results

5.1 Estimation Procedure

The specification and estimation of the econometric model(s) in the present study consist of the following stages: specification of the functional form and panel data model for the output distance function; specification of the most preferable model for technical inefficiency; and computation of technical efficiency scores from various models and comparing the results. Moreover, in order to examine whether there has been a structural change due to the introduction of environmental regulation, models are also estimated for two different regimes, i.e. non-regulated and regulated. The non-regulated regime consists of observations mainly form four plants. As noted earlier, two plants

have operated throughout study period without regulations. For the other two plants, there were no regulations in 1972-80 and 1972-82, respectively. Altogether, about 40 % of the observations in the total sample come from periods of non-regulation and 60 % from periods of regulation.

The specification of the functional form of the output distance function is somewhat problematic. In principle, a flexible functional form (e.g. translog) would be preferred to simple Cobb-Douglas functions. However, the trade-off in using a flexible functional form is the large number of parameters included, which usually causes serious multicollinearity problems. This is a problem in the present case particularly for the sub-period estimations. The non-regulated period consists of 60 observations and the regulated period of 92 observations. A full translog output distance function model with plant and period fixed effects involves 53 parameters in the present case. Experiments with this form resulted in singularity of the Hessian matrix due to multicollinearity, i.e. the data are not rich enough to allow proper identification of the parameters. Due to the above considerations, we ended up estimating a restricted translog functional form, in which the second order input terms were restricted to zero. Second order terms were used for outputs rather than for inputs, since the purpose is to derive shadow prices for outputs. Also, the cross product terms between the outputs allow for the *joint production* of multiple outputs. For the purpose of comparison, a Cobb-Douglas specification was also estimated.

In order to derive the most suitable panel data specification for the output distance function model, tests were run on whether the plant and period-specific fixed effects (dummies) should be included. According to the log-likelihood ratio tests, the best model turned out to be the two-way fixed effects model, i.e. a model with plant and period dummies included. Thus the restricted translog stochastic output distance function, shown as eq. *(5.1)*, was estimated.

$$
\begin{aligned}
(5.1)\quad \ln(D_{pt}) = {} & \alpha_0 + \alpha_p + \alpha_t + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + \beta_4 \ln(Q) + \beta_5 \ln(FLOW) \\
& + \beta_6 \ln(BOD) + \beta_7 \ln(SS) + 1/2[\ \beta_8(\ln(Q))^2 + \beta_9(\ln(FLOW))^2 + \beta_{10}(\ln(BOD))^2 \\
& + \beta_{11}(\ln(SS\ ))^2\ ] + \beta_{12} \ln(Q) \ln(FLOW) + \beta_{13} \ln(Q) \ln(BOD) + \beta_{13} \ln(Q) \ln(BOD) \\
& + \beta_{14} \ln(Q) \ln(SS) + \beta_{15} \ln(FLOW) \ln(BOD) + \beta_{16} \ln(FLOW) \ln(SS) \\
& + \beta_{17} \ln(BOD) \ln(SS) + v - u.
\end{aligned}
$$

In equation (*5.1*), α_p denotes plant-specific fixed effects (differences in plant characteristics), α_t accounts for exogenous shocks that may vary over time, but are the same for all the plants (e.g. macroeconomic factors, institutional changes, etc.). In accordance with equation *(3.5)* and the linear homogeneity of the output distance function of degree +1 in outputs, for each observation, the left-hand side and the right-hand side output variables were multiplied by $\lambda = 1 / Y_5$, where Y_5 is the output of one of the bads (nitrogen), and the second-order output terms were set to sum to zero. The Cobb-Douglas specification is obtained from (*5.1*) by setting $\beta_8 = \ldots = \beta_{17} = 0$. It should be noted that in the literature two important methodological issues have been raised concerning the estimation of a function like (*5.1*), namely, the endogeneity of the output terms on the right-hand side of the regression equation and the interpretation of the fixed effects (dummy variables) in a stochastic frontier regression equation. We now take up these issues in more detail.

In the literature there appears to be some ambiguity about the impact of the homogeneity transformation of the distance function used above on the estimation method and robustness of the results. After the transformation, the multiplicative (output) variable appears on both the left -and right-hand sides of the equations, which may result in endogeneity on the right-hand side. Atkinson, Färe and Primont (1996) have argued that this transformation necessarily leads to simultaneity bias and problems of heteroskedasticity, thus suggesting that one should use the instrumental variables estimation method with a correction for heteroskedasticity. In contrast, Coelli and Perelman for example (1996) have argued that this type of transformation does not necessarily make the right hand side variable endogenous (for a more detailed discussion on this issue, see *Section 2.2.7* of this study).

Moreover, it is not clear how one should test for endogeneity bias and heteroskedasticity in an Aigner et al. (1977) type of model, in which the error term is composed of two different parts (symmetric and asymmetric). The issue appears to become even more complicated in the case of a panel data stochastic frontier model which is estimated simultaneously with a model for determinants of inefficiency using the maximum likelihood method, as is done in the present study. Indeed, the potential heteroskedasticity problem is partly taken care of implicitly in this type of model. The stochastic frontier model includes the model for technical inefficiency, which is in fact a model for the second part of the composed error term. Since the variance of the

inefficiency term is allowed to be different for different units (e.g. pulp plants observed in different years) and is a function of a number of explanatory variables (constant, regulation, capacity utilization rate), the model does implicitly account for (explain) heteroskedasticity in the error term.

On the endogeneity issue, we took an pragmatic approach and tested for potential endogeneity bias using the Hausman specification test. Since it appears to be unclear how to test endogeneity in the stochastic frontier maximum likelihood model used in the present study, the test was computed on the OLS estimation of equation (*5.1*). More precisely, the model was estimated using both OLS and instrumental variables estimation (IV). The instruments used in the latter estimation were all the exogenous variables in equation (*5.1*) as well as the following variables that were excluded from (*5.1*): electricity consumption, wages, gross value of pulp output, nitrogen effluents, phosphorous effluents, and the production price index for the industry. The test indicated that (at the 5 % significance level) OLS coefficients are consistent (Hausman statistic at three degrees of freedom, $H_3 = 7.15$). Thus endogeneity appears not be a serious problem in the estimation.

A regards the interpretation of the plant-specific (α_p) and period-specific (α_t) effects, there also appear to be two approaches in the literature. According to the first approach, the fixed effects represent the omitted variables effects and should not be incorporated as part of inefficiency. For example, the plant-specific effects may capture effects of inputs that are invariant over time but vary across plants (e.g. type of machinery) and should not be confounded with efficiency effects. Similarly, the period-specific effects may reflect changes in the macroeconomic environment, which should not be labeled as technical efficiency effects. This type of interpretation has been adopted e.g. by Kumbhakar and Hjalmarsson (1993) and Heshamati et al. (1995). Secondly, the plant-specific and period-specific effects can be interpreted as part of the technical inefficiency which should be "added" to the inefficiency error term. This approach has been applied by Kumbhakar and Hjalmarsson (1995). In the present study, we take the first approach. Although the sulfate pulp plants form a relatively homogeneous group, they are known to use e.g. different types of capital and wood raw material, which are likely to be captured in the plant-specific fixed effects. Also, the period-specific effects capture the impact of e.g. the oil crises shocks, which should not be confounded with efficiency effects.

Equation (*5.1*) was estimated simultaneously with the model for determinants of inefficiency (Z_{pt}), i.e. equation (*5.2*) shown below.

$$(5.2) \qquad Z_{pt} = \delta_o + \delta_1 R_{pt} + \delta_2 CAPUTI_t + \omega_{pt},$$

where δ_0 is the constant term; *R* is the regulation intensity variable (see Appendix ESIII); *CAPUTI* is the capacity utilization rate of the pulp industry; and ω_{pt} is defined as above. According to Porter's hypothesis, an increase in regulatory intensity should result in a decrease in inefficiency. Since equation (*5.2*) measures inefficiency (rather than efficiency) and due to the way the *R* variable is constructed (i.e. the higher the value of *R*, the less restrictive the regulation), the results from estimating (*5.2*) can be interpreted as supporting the Porter hypothesis if the coefficient of *R* is positive. The *CAPUTI* variable is the capacity utilization rate for market pulp. It is an aggregate industry figure, rather than a plant-specific one. It measures the changes in market conditions (i.e. demand), which may constrain the ability of the plants to reach the frontier. One would expect the inefficiency to decrease as full capacity utilization is approached, and consequently the coefficient of the *CAPUTI* variable should be negative.

It may be noted that the two factors included in the inefficiency model are only a selection of potential variables which could influence the inefficiency. One general guideline for choosing the variables for the inefficiency model is that variables, which are under the direct control of the plants, should be included in the output distance function model, rather than in the inefficiency part of the model. The choice of the specific factors in the present study reflects mainly an interest in establishing the role of environmental regulation in the production process as well as data availability.

The maximization of the log-likelihood function formulated in Battese and Coelli (1993) yields consistent and efficient estimates of the parameters of the model *(5.1) - (5.2)*. The model was estimated using the *FRONTIER 4.1* program (Coelli 1994).[3]

5.2 Production Structure

We start by considering the results for the production structure and then turn to an analysis of the results for the inefficiency part of the model. *Table 5.1* gives the results

from computation of the restricted translog specification (Model 1) and the Cobb-Douglas specifications (Models 2-4). Models 3 and 4 give the results from the estimations for the sub-periods, i.e. for *regulated period* and *non-regulated period* respectively. The results for the restricted translog model for these sub-periods are not presented, since the diagnostics indicated problems with this specification. For the regulated period, the signs of the *CAPUTI* and *R* coefficients were positive, in contrast to Model 1. The mean efficiency of the model was 0.975. When the restricted translog specification was estimated for the non-regulated period, the iteration stopped at the first round. The inefficiency part of the model was not significant and the results indicated that all the observations are on the frontier, i.e. there is no inefficiency present. Further, all the input coefficients were insignificant. In summary, the large number of parameters relative to the number of observations appears to be a problem with the restricted translog model in sub-period estimations.

The results for Model 1 show that some of the input and output parameters were not significant at the 5% level. In particular, the coefficient for the labor input variable has an unexpected positive sign and is not significant. Also, the absolute values of the coefficients of the capital and materials input are unexpectedly low. The relative shadow prices computed using equation *(2.4)* are shown in *Table 5.3*. The shadow prices for FLOW and SS turned out to be positive and for BOD negative. The positive shadow prices indicate that the marginal reductions of FLOW and SS have increased plant revenues. This result for FLOW is in accordance with a priori expectations (cf. Kneese and Bower 1968) concerning the link between waste water recycling and production efficiency.

The shadow price for SS is also positive, although SS has been regulated. However, it should be noted that the shadow price relation (equation *(2.4)*) does not measure the impact of the regulation as such but rather the marginal rate of transformation between the pulp output and SS. One possible explanation for the positive shadow price is that suspended solids contain fibers, an important part of pulp, and thus recovery of waste fibers (SS) will also increase output. An other possible explanation is that the measures taken to reduce FLOW also simultaneously reduce SS, thus causing no additional costs for SS abatement. The negative shadow price for BOD indicates that lower BOD is achieved only through diverting resources from the production of pulp to the abatement process (e.g. building waste water treatment facilities).

Table 5.1 Maximum-likelihood Estimates of Stochastic Frontier Output Distance Function (for simplicity, the fixed effects are not reported)

Variable / Parameter	Rest. Translog* Model 1	Cobb-Douglas Model 2	Cobb-Douglas Model 3 R	Cobb-Douglas Model 4 No R
Constant, β_0	-10.66 (31.69)	-0.39 (0.57)	-0.44 (0.55)	-3.01 (3.02)
L, β_1	0.01 (0.40)	-0.16 (3.67)	-0.17 (4.17)	-0.08 (0.42)
K, β_2	0.03 (2.41)	-0.10 (2.81)	-0.13 (2.71)	-0.13 (1.24)
M, β_3	-0.08 (4.41)	-0.63 (12.61)	-0.58 (7.88)	-0.47 (3.91)
Q, β_4	3.40 (17.05)	0.89 (19.35)	0.95 (15.57)	0.91 (14.36)
FLOW, β_5	-0.84 (2.50)	0.08 (1.69)	0.09 (1.45)	0.25 (2.88)
BOD, β_6	0.08 (0.34)	-0.02 (1.08)	-0.04 (1.88)	-0.10 (1.74)
SS, β_7	-1.65 (9.00)	-0.001 (0.01)	-0.02 (0.93)	0.07 (2.17)
Q2, β_8	-0.17 (10.37)			
FLOW2, β_9	-0.02 (1.61)			
BOD2, β_{10}	-0.003 (0.42)			
SS2, β_{11}	-0.005 (0.58)			
QFLOW, β_{12}	0.08 (3.28)			
QBOD, β_{13}	-0.07 (3.85)			
QSS, β_{14}	0.09 (5.79)			
FLOWBOD, β_{15}	0.03 (1.65)			
FLOWSS, β_{16}	0.08 (1.50)			
BODSS, β_{17}	0.07 (7.43)			
Constant, δ_0	0.49 (7.62)	0.87 (3.42)	5.03 (4.50)	0.05 (1.08)
R, δ_1	-0.05 (2.28)	-0.84 (2.60)	-0.29 (1.42)	0
CAPUTI, δ_2	- 0.57 (7.84)	-0.72 (2.77)	-5.85 (4.50)	-0.02 (0.44)
$\gamma \equiv \sigma^2 / \sigma_s^2$	0.99 (44.29+E2)	0.58 (4.95)	0.80 (12.84)	0.99 (273.74)
Log-likelihood	355.21	175.37	118.52	94.30
mean efficiency	0.96	0.95	0.87	0.93

Table 5.2. Tests of Hypotheses for Functional Form and Parameters of the Inefficiency Models

Model	Null Hypotheses	LR-test statistic (λ)*	χ^2 critical value (5%)	Decision
functional form	H_0: $\beta_8 = \ldots = \beta_{17} = 0$	359.68	18.31	Reject H_0
M1	H_0: $\gamma = \delta_0 = \ldots = \delta_2 = 0$	23.81	9.49	Reject H_0
M1	H_0: $\delta_1 = \delta_2 = 0$	23.76	5.99	Reject H_0
M2	H_0: $\gamma = \delta_0 = \ldots = \delta_2 = 0$	11.18	9.49	Reject H_0
M2	H_0: $\delta_1 = \delta_2 = 0$	8.92	5.99	Reject H_0
M3	H_0: $\gamma = \delta_0 = \ldots = \delta_2 = 0$	28.77	9.49	Reject H_0
M3	H_0: $\delta_1 = \delta_2 = 0$	24.36	5.99	Reject H_0
M4	H_0: $\gamma = \delta_{12} = 0$	11.11	5.99	Reject H_0
M4	H_0: $\delta_1 = 0$	4.72	3.84	Reject H_0

$$\lambda = -2\left[log\left(likelihood(H_o)\right) - log\left(likelihood(H_1)\right)\right]$$

Table 5.3. Absolute Shadow Prices for Bads (evaluated at the mean of the data)

	REST. TRANSLOG (Model 1)	*COBB-DOUGLAS (Model 2)*
FLOW	340.0	90.0
BOD	-110.0	-20.0
SS	129.0	-1.0

In the Cobb-Douglas specification for the whole sample (Model 2), the input parameters have the expected negative signs and pulp output a positive sign. Also, the t-values indicate that these coefficients are significantly different from zero at the 5% significance level. The coefficient for the waste water level is positive and significant at the 10% level. The coefficients for BOD and SS are negative and not significant. The relative shadow price ratio for FLOW is 0.09, significantly lower than for Model 1.

The results from the regulated-period estimations (Model 3) are similar to the whole sample estimations (Model 2). The most significant change is that the coefficients for BOD and SS are more significant. The shadow price for BOD in terms of pulp output (-0.04) indicates that marginal increases in abatement of BOD have resulted in foregone revenues. Thus the abatement of regulated vs. non-regulated bads appear to have different impacts on plants performance. Finally, when the specification is estimated for the non-regulated period (Model 4), the results indicate positive shadow prices for FLOW and SS but negative for BOD. Moreover, the coefficients for labor and capital are not significant, even at the 10% level.

It should be noted that due to the possible multicollinearity and functional dependence between FLOW, BOD, and SS, the Cobb-Douglas estimations were also carried out using only one of the bad outputs over time. The results indicated that the absolute values of the coefficients of the bad outputs were only slightly sensitive to whether or not other bads are included. The most significant change in the results reported in *Table 5.1* was that the coefficient for *SS* was significant when estimated for the regulated period (in contrast to Model 3 in *Table 5.1*).

Although the above results suggest that reductions in FLOW have unambiguously positive spill-over impacts on plant revenues, we cannot identify the degree to which waste water reduction (recycling) has been adopted due to efficiency or environmental regulations. Probably both of these motives have played a role to some extent. Finally, it may be noted that results similar to those of the present study were obtained by Färe et al. (1993) for BOD for US pulp and paper producers in Michigan and Wisconsin.

5.3 The Determinants of Technical Inefficiency

What do the results tell us about production efficiency and the determinants of inefficiency? In *Table 5.2* the null hypothesis that there is no technical inefficiency in pulp

plants is considered for the various specifications. This test is equivalent to testing the null hypothesis that the parameters are zero, i.e. $\gamma = \delta_0 = \delta_1 = \delta_2 = 0$ or $\gamma = \delta_0 = \delta_1 = 0$. If the null hypothesis is accepted, the u_{pt} term could be omitted and the model estimated using ordinary least squares. The test results show that for each of the specifications the null hypothesis is rejected. Further, the second hypothesis considered in *Table 5.1* specifies that the inefficiency effects are not a linear function of the *R* and *CAPUTI* variables ($\gamma = \delta_1 = \delta_2 = 0$ or $\gamma = \delta_1 = 0$). The null hypothesis is rejected for all three models at the 5% level of significance. In summary, these tests indicate that both the *R* and *CAPUTI* variables should be included in all the models (naturally, *R* is not included in Model 4, which is estimated for the period in which no regulation was imposed).[4]

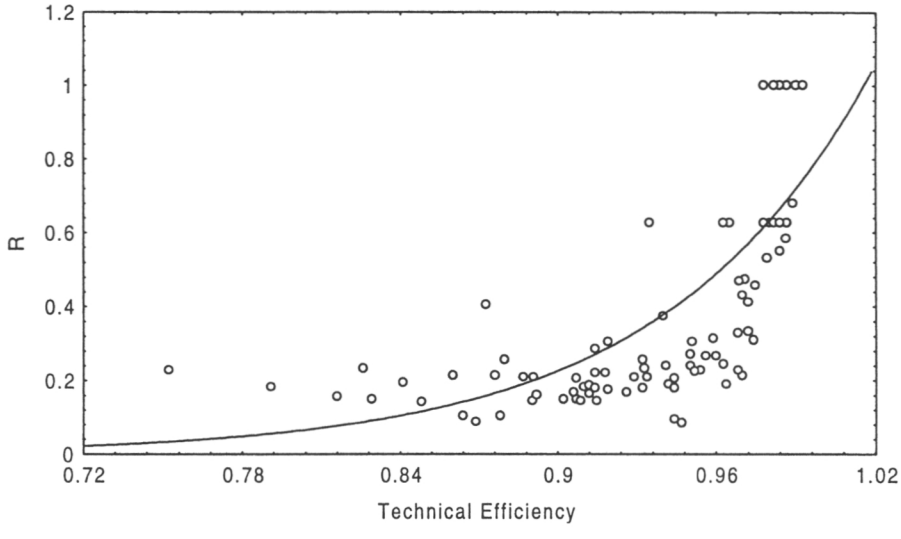
As Coelli (1995) has indicated, the interpretation of the γ parameter is not as clear in the above specification as it is in the conventional half-normal stochastic frontier model in which the determinants of inefficiency variables are not included (i.e. all δ_i are zero). According to Coelli (1995), the γ parameter may be interpreted loosely in the present context as an indication of the amount of unexplained variation in the technical inefficiency effects, relative to the sum of this value and the variance of v_{pt}. The parameter γ gets values between zero and one. If γ is zero, then the variance of the inefficiency effects is zero and the model reduces to a traditional mean response function. On the other hand, a high value for the γ parameter indicates that the model of determinants of inefficiency accounts for the bulk of the variation in the technical inefficiency. In all the model specifications shown in *Table 5.1*, the absolute value of γ parameter is between 0.58-0.99 and is statistically significant at the 5% level.

The signs of the δ_i coefficients are of particular interest.[5] A negative sign for the estimated coefficient of regulatory intensity variable (*R*) indicates that an increase in the value of the regulatory variable, i.e. a decrease in the intensity of regulation, will result in a decrease in the value of the technical inefficiency effect. Thus the more restrictive the regulation, the more inefficient the production process will be. This result is contrary to one of the central arguments in Porter's hypothesis. This result holds for the whole estimation period as well as for the regulated period alone.

In *Figure 2*, the relationship between the environmental regulatory intensity variable (*R*) and technical efficiency (computed from Model 2) is illustrated in the simple scatterplot. The more restrictive the regulation (i.e. the lower the value of *R*), the less

efficient the plant tends to be. In fact, there seems to be a sharp dividing line, with regulation having very little impact until about $R = 0.7$, below which there is immediately a much greater reduction and variation in efficiency. This observation appears to contradict with Porter and van der Linde (1995), who argue that stringent regulation can produce greater innovation and innovation offsets than lax regulation (see the quote in p. 128 of this study). Instead *Figure 2* suggests that plants facing either no regulation or lax regulation do much better than those facing stringent regulation.

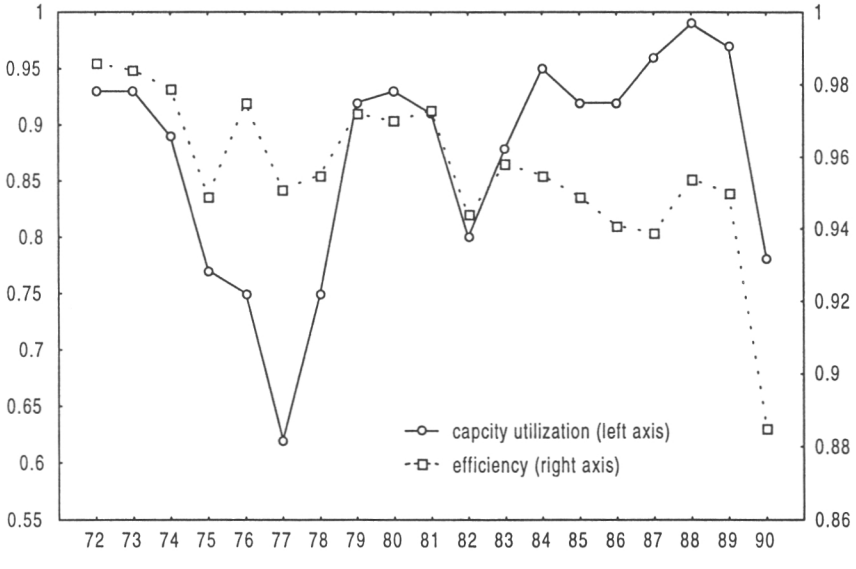
Figure 2. Scatterplot of Regulation Intensity (R) and Technical Efficiency, 1972-90 (from Model 2)



The coefficient of the capacity utilization rate factor is negative, as one would expect, in all models. Thus the higher the capacity (or the better the external market conditions), the more efficient the plants. In *Figure 3.* the changes in efficiency scores computed from Model 2 and the changes in the capacity utilization rate are shown over time. The figure shows that technical efficiency has decreased over time. Whether this is a result of tighter regulations, ageing of fixed capital investments (plant and machinery), or some other reason, remains unclear. The figure also indicates that there is a strong link between capacity utilization and production efficiency over time. In particular, low capacity utilization rates are associated with low efficiency, and vice versa. In other

words, technical inefficiency tends to be greater during economic slumps. Thus it appears that plants adjustment to economic cycles and market conditions takes place partly through technical efficiency. Therefore, adjustment not only takes place on the production function but partly by moving below the production frontier. This is not surprising in light of the fact that it is a common practice for pulp producers to shut down their operations during periods of weak market demand, thereby idling both capital and labor. (Although during the shutdowns, the machines are not running, labor time does not necessarily decrease accordingly. Typically, for example, maintenance work is done during shutdowns.).

Figure 3. Capacity Utilization Rate and Technical Efficiency, 1972-90 (from Model 2)



5.4 Technical Efficiency Scores

Given equations (*5.1*) and (*5.2*), the technical efficiency of production of the *p*-th plant in the *t*-th year is defined by equation (*3.6*). The technical efficiency scores rely upon the value of the unobservable u_{pt} predicted. The details of obtaining the values for the conditional expectation *of* $exp(-u_{pt})$, given the value of $e_{pt} = v_{pt} - u_{pt}$, is described in

Battese and Coelli (1993). Descriptive statistics for the technical efficiency scores of the different models are given in *Table 5.4.*

Table 5.4 Technical Efficiency Scores for Different Models and Plant Specific Scores for Model 2

MODEL	MEAN	MIN	MAX	STD.DEV
M1	0.96	0.84	1	0.03
M2	0.95	0.75	0.99	0.05
M3	0.87	0.25	0.99	0.19
M4	0.93	0.73	1	0.06
PLANT (M2)				
plant 1	0.97	0.84	0.99	0.03
plant 2	0.92	0.79	0.99	0.05
plant 3	0.96	0.88	0.99	0.03
plant 4 (no regu.)	0.99	0.98	0.99	0.002
plant 5	0.96	0.83	0.99	0.04
plant 6 (no reg.)	0.98	0.97	0.99	0.01
plant 7	0.92	0.75	0.99	0.06
plant 8	0.92	0.86	0.98	0.04

The results *Table 5.4.* show that the functional specification does not appear to have a large impact on efficiency scores; the mean values for the restricted translog and Cobb-Douglas models are 0.96 and 0.95 respectively. However, the efficiency scores do change depending on whether the model is estimated for the regulated or non-regulated period or for both periods simultaneously. The regulated period mean efficiency (0.87) is considerably lower than for the whole sample or for the non-regulated period (0.93). The results from Model 2 indicate that on average plants produce 95% of the output that could be theoretically produced with the same bundle of inputs by a technically efficient plant. When the means of the technical efficiency scores are compared across plants (see the bottom panel of *Table 5.4*), the result indicates that plants with no regulation (plants 4 and 6) are the most efficient. The minimum plant-specific mean value is 0.92 and the maximum 0.99. Thus the most efficient plant is on average 7% more efficient than the least efficient. The results also indicate that the plants have been more efficient during the non-regulated period (Model 4) than in the regulated period (Model 3).

It should be noted that the above results are in line with the findings of Gray and Shadbegian (1993) for US pulp and paper plants. Their findings are also based on plant-level panel data and indicate that the more regulated pulp and paper plants have significantly lower productivity levels and slower productivity growth than less regulated plants.

In summary, both the results from the model for determinants of inefficiency and the technical efficiency scores indicate that environmental regulation has had a negative impact on the production efficiency of the pulp plants. Consequently, the results do not support the central premise of Porter's hypothesis. On the other hand, the internal process changes (recirculation of waste water) which were adopted mainly to increase production efficiency, have reduced water pollution significantly. However, as Kneese and Bower (1968) have observed, the causal link appears to run from production efficiency to abatement, not vice versa, as Porter's hypothesis would suggest. Indeed, these considerations beg the question: To what extent do the evidence and case studies cited in support of Porter's hypothesis (e.g. in Porter and van der Linden 1995) reflect the fact that, besides the regulation, measures taken to increase production efficiency may also lead to a reduction of bads. In other words, independent of the regulations, firms may adopt pollution prevention measures when they are associated with increases in production efficiency.

6. Conclusions

The role of environmental regulations as a source of production efficiency and competitiveness of firms has been raised by Porter's influential but still controversial hypothesis. The discussion of Porter's hypothesis has been based to a large extent on theoretical considerations and anecdotal references to case studies (Porter and van der Linde 1995, Palmer, Oates and Portney 1995). The present study is an attempt to estimate explicitly the impact of environmental regulation on production efficiency in the pulp industry. Furthermore, shadow prices were derived for regulated and non-regulated bad outputs. The framework used to obtain the results extended the existing literature on stochastic frontier functions to include undesirable outputs and environmental regulation.

The results show that an increase in regulative intensity has led to a decrease in the production efficiency of pulp plants. The results, therefore, do not appear to give

support to a central premise of Porter's hypothesis but instead agree with the simple economic textbook model, whereby an increase in the stringency of environmental regulations unambiguously makes the polluting firm worse off. Thus the environmental standard on BOD (and suspended solids) as such has a negative impact on production efficiency. However, the positive shadow price for the unregulated bad output (FLOW) indicated that reduction of bads can also be profitable for the plants. This result can be interpreted to be in accordance with the statement by Porter and van der Linde (1995, p. 105) that *"Fundamentally, it (pollution) is a manifestation of economic waste and involves unnecessary, inefficient or incomplete utilization of resources, or resources not used to generate their highest value."* But unlike Porter and van der Linde appear to suggest, the driving motivation behind the reduction of waste water may have been the aim to increase production efficiency rather than the environmental regulation per se. Indeed, there has been no regulation of FLOW. Moreover, in the case of the present study, two plants operated throughout the study period without regulations but still reduced water effluents to the same degree as the regulated plants.

At a more practical level, the results indicate that the water quality authorities may have failed to design regulations that promote the use of innovative measures through which plants could have realized overlooked opportunities to increase efficiency. Indeed, the above results suggest that the regulator should directly control the FLOW parameter instead of BOD and suspended solids. In other words, regulation should be targeted at pollution-prevention measures (closing the water system), rather than end-of-pipe measures (e.g. biological treatment lagoons). However, in relation to the Porter hypothesis, the positive shadow price of FLOW also raises the question that if the water authority had regulated FLOW instead of BOD and SS, would the impact of the regulation on production efficiency have had been positive rather than negative, as the Porter hypothesis suggests.

Notes

1. Grosskopf and Hayes (1993) use the methodology to examine the local public sector bureaucrats input choices.

2. Kneese and Bower (1968) state that *"Changes in production processes, however, have the major impact on wastes generation, although it should be emphasised that technological changes have not been, and are not usually, instituted because of water quality problems. In fact, most changes in production technology have been stimulated by factors unrelated to water problems, and have been developed without explicit considerations for their effects on water quality" (pp. 44-45).*

3. The *FRONTIER 4.1* program uses a three-step procedure in producing the maximum likelihood estimates of the parameters of a stochastic frontier function. First, ordinary least square (OLS) estimates of the function are obtained. In the second stage, a two-phase grid search of a variance term (i.e. a term that measures the ratio between the variance of the inefficiency component relative to the total variance of the residual) is conducted. The right hand side parameters (except the constant term) are set to the OLS values. The variance term can take values between 0 and 1 and the grid search considers values from 0.1 to 0.9 in increments of size 0.1. Finally, the values selected in the grid search are used as starting values in an iterative procedure (using the Davidon-Fletcher-Powell Quasi-Newton method) to obtain the final maximum likelihood estimates.

4. It may be noted that we experimented with different lead indicators for the *R* variable. Since the pulp plants usually know approximately 1-5 years in advance of the regulation, that is going to be imposed on them, the plants may anticipate and take actions already before the regulation comes in to force. However, the results showed that the computations with 1,2,3, and 4 year lead indicators did not perform better than the current period regulation intensity variable.

5. Coelli (1995) has noted that the interpretation of the absolute magnitudes of the δ_i coefficients is not straightforward. He suggests that one informative way to interpret the sizes of these coefficients would be to obtain the partial derivatives of the technical efficiency predictor with respect to each of the plant-specific factors. However, since the main interest in the present study is in the sign of the δ_i coefficients, rather than their absolute magnitude, these are not presented here. Still, it may be noted that values of the δ_i coefficients relate to the way the plant-specific factors shift the mean of the (pre-truncated) distribution of u_{pt}. The size of the δ_i coefficients are probably smallest in the restricted translog model, since this model leaves the least to be explained in the error term.

References

Aigner, D., C.A.K. Lovell and P. Stemmed (1977). Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, pp. 21-37.

Atkinson, S., R. Färe, and D. Primont (1996). Measuring Technical and Allocative Inefficiency via Econometric Estimation of Distance Functions, unpublished mimeo; May 20, 1996.

Battese, G.E: & T.J. Coelli. (1993). A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects, Working Papers in Econometrics and Applied Statistics No. 69, Department of Econometrics, University of New England, Australia.

Battese, G.E: & T.J. Coelli. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, to be published in *Empirical Economics, vol. 20, no. 2*.

Bernstein, MA, S.L. Feldman, and A.P. Schinnar (1990). Impact of pollution controls on the productivity of coal-fired power plants, *Energy Economics*, pp. 11-17.

Brännlund, R., R. Färe, and S. Grosskopf (1995). Environmental Regulation and Profitability: An Application to Swedish Paper and Pulp Mills, *Environmental and Resource Economics* 5; 1-14.

Brännlund, R. and K.-G. Löfgren (1996). Emission Standards and Stochastic Waste Load, *Land Economics*; 218-30.

Coelli, T. (1994). A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, Department of Econometrics, University of New England, mimeo.

Coelli, T. (1995). Measurement and Sources of Technical Inefficiency in Australian Coal-Fired Electricity Generation, paper presented at the New England Conference on Efficiency and Productivity, University of New England, Armidale, November 23-24, 1995.

Coelli, T. and S. Perelman (1996). Efficiency measurement, multiple-output technologies and distance functions: with application to European Railways. CREPP Working Papers 96/05, Universite de Liege.

Coggins, J.S. and J. R. Swinton (1996). The Price of Pollution: A Dual Approach to Valuing SO2 Allowances, *Journal of Environmental Economics and Management* 30: 58-72..

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

Gollop, F.M. and M. J. Roberts (1983). Environmental Regulations and Productivity Growth: The Case of Fossil-fueled Electric Power Generation, *Journal of Political Economy*, vol 91.; 654-673.

Gray, W.B. & S.J. Shadbegian (1993). Environmental Regulation and Manufacturing Productivity at the Plant Level, National Bureau of Economic Research, Working Paper No. 4321.

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

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

Grosskopf, S., K. Hayes, L. Taylor and W. Weber (1996). Budget Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling, *Review of Economics and Statistics*, forthcoming.

Heshmati, A., S. C. Kumbhakar, and L. Hjalmarsson (1995). Efficiency of the Sewdish Pork Industry: A farm level study using rotating panel data 1976-1988, *European Journal of Operational Research*; 519-534.

Hetemäki, L. (1995). Using Deterministic and Stochastic Distance Functions to Measure the Effects of Pollution Control on a Firm, Finnish Forest Research Institute, Research Papers no. 549.

Jaffe, A. B. and K. Palmer (1996). Environmental Regulation and Innovation: A Panel Data Study, National Bureau of Economic Research, Working Paper 5545, April 1996.

Jaffe, A.B., S.R. Peterson, P.R. Portney & R.N. Stavins (1995). Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?, *Journal of Economic Literature*; pp. 132-163.

Kneese, A.V. and B.T. Bower (1968). *Managing Water Quality: Economics, Technology, Institutions*. Resources for the Future, John Hopkins Press.

Kumbhakar, S. B. and L. Hjalmarsson (1993). Technical Efficiency and Technical Progress in Swedish Dairy Farms, in Fried, H.O., C.A.K. Lovell & S.S. Schmidt (eds.).

Kumbhakar, S. B. and L. Hjalmarsson (1995). Labour-Use Efficiency in Swedish Social Insurance Offices, *Journal of Applied Econometrics*, vol. 10; 33-47.

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.

Palmer, K., W. E. Oates, and P. Portney (1995). Tightening Environmental Standards: The Benefit-Cost or the No-Cost Paradigm?, *Journal of Economic Perspectives*, vol 9, No. 4; 119-132.

Pittman, R.W. (1981). Issues in Pollution Interplant Cost Differences and Economies of Scale, *Land Economics*, 1-17.

Pittman, R.W. (1983). Multilateral Productivity Comparisons with Undesirable Outputs, *Economic Journal,* 883-891.

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

Porter, M.E. (1991). America's Green Strategy, *Scientific America*, April, 264: 168.

Porter, M. E. and C. van der Linde (1995). Toward a New Conception of the Environment-Competitiveness Relationship, *Journal of Economic Perspectives*, vol 9, No. 4; 97-118.

Simpson, R.D. and R. L. Bradford (1996). Taxing Variable Cost: Environmental Regulation as Industrial Policy, *Journal of Environmental Economics and Management*; 282-300.

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

Yaisawarng, S. & Klein, D.J. (1994). The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry, *Review of Economics and Statistics*; 447-460.

Appendix ES. III. Regulation Intensity Variable

Similarly to Gollop and Roberts (1983), a variable *R*, which describes the regulatory intensity is constructed. Gollop and Roberts construct a regulatory intensity variable, which gets values between zero and unity. The variable is unity if no environmental regulation is imposed on the firm. The lower bound zero is obtained if the legal standard is greater than or equal to the firm's actual effluent rate, i.e. this would imply that the standard is of no importance to the firm.

The Gollop and Roberts intensity measure has the drawback that it does not take account of the fact that the standard is seldom satisfied exactly in reality, at least in the pulp industry. The plants are usually operating under the legal standard level, and sometimes even above it due to stochastic shocks, such as equipment failures, failures in process control, and human failures (Brännlund and Löfgren 1996). At least of some importance is the fact that the plants usually receive the content of the new standard 2 - 5 years before they come into force. Consequently, the plants may take measures before the standard is actually implemented. Thus, their effluents may also be significantly lower than the current standard would allow. Thus, the fact that the legal standard is often higher than actual effluents does not necessarily imply that the regulation is ineffective.

According to water authority officials, a good parameter describing the intensity of the regulation is the standard/output ratio. That is, the number of tons of *SS* and/or *BOD* the plants are allowed to discharge per ton of pulp produced. The way the *R* variable is constructed in the present study reflects this concern. The higher its value, the less restrictive the regulation standard.

Thus, the ratio of the discharge standard (tons) to pulp produced (tons) is calculated as

$$(A1)\ RSS_{pt} = \frac{SSS_{pt}}{Q_{pt}}; \quad RBOD_{pt} = \frac{SBOD_{pt}}{Q_{pt}},$$

where SSS is the discharge standard for suspended solids (tons), *SBOD* is the discharge standard for biological oxygen demand (tons), and Q is pulp output (tons). For the plants and periods for which no standard was imposed (60 observations out of the total of 152 observations), the values of *RSS* and *RBOD* are set to 10 per cent higher than the maximum values obtained in (A1). Thus, it is assumed that the least restrictive regulation imposed during the study period, increases the intensity of the regulation by 10 per cent compared to the case with no regulation. Although this assumption is arbitrary, the results are not sensitive to small changes in the maximum value.

In line with Gollop and Roberts intensity variable, the *R* variable is formed so that it gets values between zero and one $(0 < R \leq 1)$. This is obtained by scaling the *RSS* and *RBOD* series by their respective maximum values. Moreover, we construct the *R* variable as a weighted average of the two effluent standards. Thus the two standards are assumed to receive equally strong weight in the final regulation intensity variable. To summarize, the regulation intensity variable, *R*, has an upper bound of unity when the water authority has mandated no effluent standard for *SS* and *BOD*. The closer *R* is to zero, the more restrictive the regulation.

Appedix I. Institutional Setting for Pulp Production and Effluent Control

Pulp Processes

The production of wood pulp across the globe rests basically on two different processes; alkaline *sulphate* (Kraft or chemical) and *thermocehmical* (TMP or mechanical) pulping. In the kraft pulping process, used to make most chemical pulp, a solution of sodium hydroxide and sodium sulfide dissolves the nonfibrous materials. In mechanical pulping, the fibers are separated from each other by refining or grinding. Moreover, techniques to minimize wastes differ as between the mechanical and sulfate pulping processes. Consequently, if one wants to analyze the impact of effluent control on pulp plants, it is clear that these two processes should be separated from each other in order to obtain accurate results. Indeed, Pittman (1981), who studied the impact of environmental regulation on different pulping processes, notes that, "*The different pulping processes represented in the sample -and the associated differences in end-product characteristics- are troublesome for purposes of estimation; technological homogeneity is desirable in production function analysis. In particular, Bower et al. (1971), Bower (1975), and Krutilla and Smith (1979) have noted the importance of end product characteristics for pollution control requirements in this industry (footnote 7, p.4).*" In the present study only the sulfate pulp process is studied. In terms of both economic and environmental impact, the sulfate pulp process is clearly of more importance than mechanical pulp in Finland.

In the sulfate pulp process, about half of the wood material, primarily lignin, dissolves into the cooking liquor. This dissolved matter is recovered and burned, and some of the energy released is used to run the process. The sulfate pulp process itself essentially produces all the energy it needs. The yield of fibers is about 50 % of the original weight of the wood. It should be noted that the sulfate pulp processes can in turn be divided into different sub-processes depending on the bleaching and delignification technology adapted. However, the present analysis does not analyze separately these sub-processes, because the number of observations in the analysis (estimations) would have to be reduced significantly.

Sulfate Pulp Production in Finland

The sulfate 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 sulfate pulp (for export or sale to domestic paper plants). In 1990 there were altogether 17 sulfate pulp plants in Finland, of which 7 were non-integrated. The sulfate 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.

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 sulfate pulp plants over the period 1972-90. All the plants are non-integrated, except one, for which it was possible to separate the sulfate pulp production from the paper/paperboard production in the data. The plants in the sample accounted for more than a half of the total production of the sulfate pulp industry during 1972-90.

Pulp Plant Effluents

Effluent from pulp mills is composed of three major types of constituents: dissolved organic compounds, suspended solids, and an inorganic component. Each of these effluent components can degrade water quality and aquatic habitat in rivers, estuaries and coastal waters. Dissolved organic material in the effluent breaks down as it is digested by bacteria (either aerobic or anaerobic). The former places extra demands on dissolved oxygen for aerobic decomposition in receiving waters. Typically, the different effluent loads are described by using various parameters, such as *biological oxygen*

demand (BOD), suspended solids (SS), nitrogen (N), phosphorus (P), chemical oxygen demand (COD), and absorbable organic chlorine (AOX). In the present study, our interest is in BOD and SS, which are the parameters the water authority has regulated.

BOD is a parameter which describes the amount of oxygen the micro-organism uses within a specific time period (usually 7 days demand) for biologically dissolving waste water, and it gives an indication of the amount of easily dissolvable matters in the waste water. BOD can deplete the dissolved oxygen required by fish and other aquatic species. Under such conditions, anaerobic decomposition is likely to take place, with a resultant release of noxious gases, such as methane, hydrogen sulphide and ammonia. If the amount of oxygen changes in the receiving water, it affects the fish population. The suspended solids (SS) component of pulp mill effluent, which is primarily made up of cellulose fibers, wood particles and small pieces of bark, and lime mud, can settle out of suspension near mill outflows, smothering the benthic habitat of various species and the fish population in general. SS can also create mud-banks which change the structure of the waterbed and thereby affect fish and other creatures in the receiving waters.

Effluent Regulations

Up until the beginning of the 1970s water pollution control was not a very serious problem for the Finnish pulp industry because little investment was required to control waste water. However, public concern over the quality of water resources and the fact that the Finnish pulp and paper industry was the main polluter of water resulted in tighter control measures. The Finnish Water Act was drafted in 1962 and the first waste water permits were issued at the end of the 1960s. The regulation of effluent loading is based to a large extent on discharge permits and Water Rights Court procedures, both of which differ from common practice in most other countries. This is mainly because most waters in Finland are in private hands. Thus, for example, a fishing community owning the property rights to waters near the pulp plants have been able to obtain compensation payments from the plants due to the reduction in the quality of their waters. The local water authority regulates effluents by specifying absolute effluent limits for each individual plant. According to the water law (VL 10:24), a permit is issued for effluents that cannot be eliminated at *reasonable* cost. The authority has to weigh the economic benefits of the firm against the costs to the environment. Thus, acceptable standards are based on

individual judgments and often on compromise.

Every plant has its own specific regulation on the substances that it discharges. The local water authority specifies effluent limits (standards) for BOD and SS as maximum allowed discharge, expressed in total load per time unit. In some cases limits are also given as specific load per ton of product or as concentrations and removal efficiencies. In general, the limits must be attained as mean values for 1,3 or 6 months, depending on the size and type of plants. Permits are typically renewed every 3-10 years. Permits have to be renewed typically after 3-10 years. In addition to standards, the regulatory authorities have controlled waste water effluents by giving tax credits for investment in pollution-control equipment and by ordering the plants to build special devices to clean the waters. However, the two latter regulation measures have had only a relatively small impact compared to the other measures (tax credits have amounted to a very small proportion of total investments in abatement capital). Occasionally the water authority has required the plants to take specific measures to control effluents, such as building external waste water treatment facilities.

In practice, the regulations have been very heterogeneous across the plants. Documents from the water authorities indicate that it is difficult to find any clear evidence which would show that a common policy has been applied simultaneously to all plants. Also, two plants in the study were never subject to regulation during the study period (1972-1990), and the remaining six plants were at times in 1970s not subject to regulation.

The water authority carries out water pollution monitoring in order to assess the waste water ingredients, their quantity and toxicity; to control 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 (local water authority). The monitoring is carried out by both the official water laboratory (of which there are about 20 in the whole country) and by the pulp plants themselves.

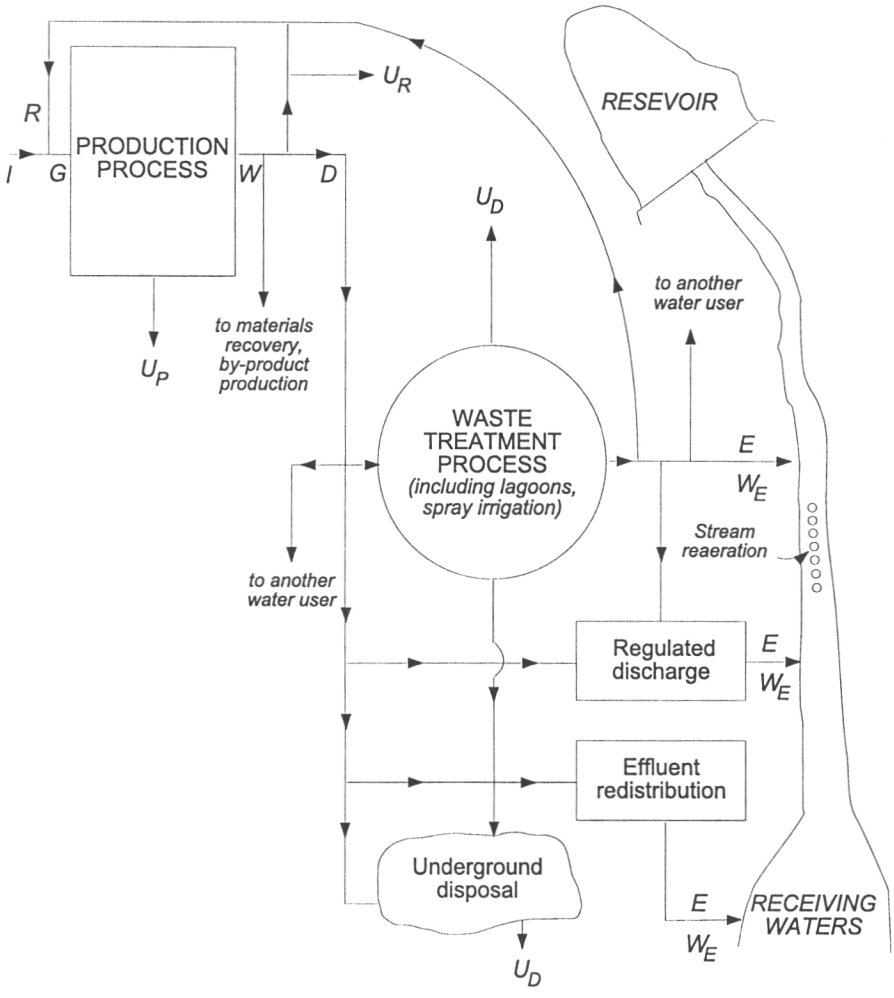
Effluent Management in Pulp Plants

There are basically two different approaches to effluent control in pulp plants: first, by internal process changes (pollution-prevention methods); and second, by external treatment methods (end-of-pipe methods).Within these two categories a number of sub-

techniques are available. Internal process changes reduce the amount of total waste water flow (FLOW) discharged to receiving waters, thus also reducing the noxious substances, such as BOD and SS, in the waste water. Typically, these internal process change measures are called “closed-loop water systems”. The external treatment methods can be divided into four different categories: 1) biological, 2) chemical, 3) mechanical, and 4) physical. The most important (and costly) external treatment method has been the biological treatment. The purpose of biological waste water treatment is to convert dissolved and colloidal pollutants to cell mass, water, and gases. Biological treatment in turn can be of three types: aerated lagoon, activated sludge units, and trickling filters. In the *Figure*, below, a simple diagram illustrating the effluent control of a pulp plant is provided. The *Figure* is based on Kneese and Bower (1968, p. 43), where a more detailed description of the water utilization, waste generation, and waste disposal of pulp and paper plants is presented.

It is important to note, that it is difficult to separate the process changes which have been adapted due to environmental considerations from those adapted purely (or partly) due to increasing the efficiency of the production process in general. For example, the sulfate pulp plants no longer discharge their "black liquor" waste resulting from wood pulp fiber separation to surface waters. They now dehydrate this waste and use wood bark wastes and sawdust to fuel a boiler that generates electricity. They are also able to recover molten inorganic chemicals and recycle "black liquor" back into virgin "white liquor" used to separate wood fibers. This chemical recovery cycle has a substantial impact on both the economics of pulp production and a positive impact on the environment. Recovery means capturing the spent cooking chemicals and waste fibers, processing them, and chemically converting them to a reusable form. Thus, the sulfate pulp chemical and waste water recovery process reduces chemical and wood costs by constantly reusing the same chemicals and by recovery of fibers. Simultaneously, this recovery also reduces adverse environmental impacts.

*Figure A. Pulp Plant Water Utilization, Waste Generation, and Waste Disposal**



I = Intake water.
R = Recirculated water.
G = Gross water applied for all in-plant uses.
U = Consumptive use or net depletion of water. $U = U_P + U_D + U_R$, where U_P = consumptive use in the production process, U_D = consumptive use in wastewater disposal system, and U_R = consumptive use in the recirculation system.
D = Wastewater discharge from the production process.
E = Final effluent from the production unit (available for reuse). Where a lagoon or spray irrigation system is involved, the final effluent (if any) consists of seepage, lagoon discharge, and/or surface runoff.
W = Waste load generated in all operations of the production process, e.g. pounds of BOD, Btu's of heat.
W_E =Waste load on the final effluent, i.e. pounds of BOD, Btu's.
Degree of recirculation = R/G X 100%.

**Figure adapted from Kneese and Bower (1968).*

References

Bower, B. (1975). Studies of Residuals Management in Industry, In Economic Analysis of Environmental Problems, ed. Mills, E.S. New York: National Bureau of Economic Research.

Bower, B. , G. Lof and W.M. Hearon (1971). Residuals Management in the Pulp and Paper Industry, Natural Resources Journal (Oct.); 605-23.

Kneese, A.V. and B.T. Bower (1968). Managing Water Quality: Economics, Technology, Institutions. Resources for the Future, John Hopkins Press.

Krutilla, J.V. and V. K. Smith (eds.1979). The Implications of Potential Resource and Environmental Constraints for Economic Growth. NSF-C-ERS77-15083. Washington D.C.: Resources for the Future/National Science Foundation.

Pittman, R. (1981). Issue in Pollution Control: Interplant Cost Differences and Economies of Scale, Land Economics; 1-17.

Appendix II. 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 types of statistics are based on annual questionnaires sent to all plants. Due to the fact that the questionnaires were sent by two different authorities for the collection of different information, the data were not always coherent. Consequently, 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 firm, the plants are designated only by code numbers.

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

Water Pollution (FLOW, BOD, SS). The water pollution statistics used in the present study include the flow of waste water (FLOW) (m3/a), biological oxygen demand (BOD) (t/a) and suspended solids (SS) (t/a) /Source: National Board of Waters and Environment). The data on water pollution regulations (standards on BOD and SS during the study period) are also from the National Board of Waters and Environment. It may be noted, that the water pollution statistics concerning the quantities of effluents of the pulp plants is, in general, considered to be of good quality by the National Board of Waters and Environment (see Enckell-Sarkola, Pitkänen and Valve 1988). However, some of the figures were inconsistent and so they were checked and corrected by directly contacting the plants involved.

Labor (L). The data on labor input cover 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 (1979). The quantity of labor input is measured as hours worked. Since there may be differences between production and non-production workers that are 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 (K). 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 capital series comprise annual (1974-90)

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 (1973) was used. According to this procedure the depreciation rate is calculated as $(1-\delta)L = X$, 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 (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 (M). The data on intermediate materials comprise the value of materials (including electricity). 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 are 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.

Capacity Utilization Rate (CAPUTI). The data is industry-level capacity utilization rate for market pulp. It can be considered to be a good approximation of the market conditions in the sulfate pulp industry. For years 1980-90 it was obtained from the *Yearbook of Forest Statistics, Finnish Forest Research Institute*; and for years 1972-79 from the pulp industry marketing organization, *Finncell.*

The descriptive data are given in Table A2. The standard deviations for all the variables are less than their mean values, indicating that the mills are a relatively homogeneous group.

Table A2. Descriptive Statistics. 8 Sulfate Pulp Plants Observed Annually During 1972-90 (sample size = 152)

VARIABLE	UNIT	MEAN	ST. DEV	MIN	MAX
Q	1000 t	224.9	94.3	87.3	511.8
GVPQ	FIM/ton	2729.2	373.7	1697.4	3833.1
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
FLOW	mill. m3	42.1	20.4	15.2	126.7
BOD	1000 t	6034.8	2988	554	15370
SS	1000 t	1978.3	1569.3	277	9950
CAPUTI	%	0.87	0.10	0.62	0.99

Q = pulp output (tons); GVPQ = 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; FLOW = waste water flow (cubic meters); BOD = biological oxygen demand (tons); SS = suspended solids (tons); CAPUTI = capacity utilization rate of the pulp industry.

References

Berndt, E. (1991). *The Practice of Econometrics*, Addison-Wesley

Enckell-Sarkola, E. , H. Pitkänen and M. Valve (1988). Pollution Load Monitoring in Finland. Paper presented at the Pollution Load Monitoring Symposium 5-9 April 1988, Tallinn.

Kuh, E. and R. Schmalense (1973). *An Introduction to Applied Macroeconomics*, North-Holland.

Mäisti, E. (1979). Tulonjako paperiteollisuudessa vuosina 1955-1977, Työväen Taloudellinen Tutkimuslaitos, tutkimusselosteita 8:1979.

Nickell, S., S. Wadhwani, and M. Wall. (1992). Productivity growth in UK companies, 1975-1986, European Economic Review 36; 1055-1091.

Simula, M. (1979). Tuottavuus Suomen metsäteollisuudessa. Licentiate thesis, University of Helsinki, Department of Social Economics of Forestry.

Appendix III. Panel Data Models

In the present study, it has been assumed that the slope coefficients are constant across plants, and differences between plants are captured through intercept or disturbance terms. The five different model specifications, outlined below, were used to estimate the econometric model (see Hsiao 1986).

The pooled cross-section time-series model (eq. 1), in which the constant terms are the same across plants (i.e., only one common constant term, α_0), was estimated using Ordinary Least Squares (OLS).

$$(1) \qquad D_{0pt} = \alpha_0 + \beta' X_{pt} + \varepsilon_{pt},$$

In equation 2, it is assumed that differences across plants can be captured in differences in the constant term. Thus, α_p is an unknown parameter to be estimated using dummy variables (d_p) indicating the *p*th plant. This model is the Fixed Effects (FE) model and it is estimated using OLS.

$$(2) \qquad D_{0pt} = \alpha_p + \beta' X_{pt} + \varepsilon_{pt}.$$

The FE model can be extended to include a time-specific effect as well, i.e. it is assumed that similar effects "hit" every plant in each particular period. The extended model is known as the Two Factor Fixed Effects (TFFE) model. This model is obtained by the addition of an additional T-1 dummy variable to equation (2); it is also estimated by OLS.

$$(3) \qquad D_{0pt} = \alpha_0 + \alpha_p + \lambda_t + \beta' X_{pt} + \varepsilon_{pt}.$$

Often the fixed effects models are viewed as applying to the cross-sectional units in the study, not to additional ones outside the sample. Since the present sample of sulfate pulp plants is not an exhaustive sample (there are between 12 and 17 plants altogether, depending on the definition and time period one is using); it may be more appropriate to model plant-specific constant terms as randomly distributed across cross-sectional plants. The Random Effects (RE) model is shown in (4), where the term μ_p is the random disturbance characterizing the *p*-th observation and is constant through time. In other

words, the assumption that the μ_p are random variables implies that the P plants can be regarded as a random sample from some larger population and that the μ_p and X_p are uncorrelated. The efficient estimator for model (4) is Generalized Least Squares (GLS). First, the variance components are estimated using the residuals from the OLS regression. In the second stage, feasible GLS estimates are computed using the estimated variances.

$$(4) \qquad D_{0pt} = \alpha_0 + \beta' X_{pt} + \mu_p + \varepsilon_{pt}.$$

It is possible that the random effects also vary through time and that the correct model is the Two Factor Random Effects (TFREM) model (eq. (5)). The term v_t captures the time varying random effects. The TFREM model is estimated consistently and efficiently by feasible GLS.

$$(5) \qquad D_{0pt} = \alpha_0 + \beta' X_{pt} + \mu_p + v_t + \varepsilon_{pt}.$$

Panel Data Specification Test

The Hausman test (m) is a chi-square test of whether the generalized least squares (GLS) estimator is an appropriate alternative to the fixed-effects estimator. If the individual effects are not correlated with the other regressors, OLS and GLS are consistent but OLS is inefficient. In the opposite case, OLS is consistent but GLS is not. It should be noted that when P is fixed and T tends to infinity, β_{FE} and β_{GLS} become identical. Consequently, the Hausman test,

$$m = \hat{q}' \hat{V}ar(\hat{q})^{-1} \hat{q}, \quad \text{where } \hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}, \hat{V}ar(\hat{q}) = Var(\beta_{FE}) - \hat{Var}(\beta_{GLS}),$$

approaches zero. Thus, the Hausman test cannot be used to test for misspecification. However, in this case the fixed-effects and random-effects models are indistinguishable for all practical purposes (Hsiao 1986). It should also be stressed that when P is fixed and $T \to \infty$, the maximum likelihood estimates of μ, β, and σ_μ^2 converge to the fixed-

effects estimator and are consistent, but the MLE of the variance of the plant effects is inconsistent. This is due to the fact that when P is fixed, there is not enough variation in the plant effects no matter how large T is (Hsiao 1986).

References

Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge University Press.

Appendix IV. Test Statistic

For a detailed description of the different test, see e.g. Godfrey (1988) and Davidson and MacKinnon 1993).

Autocorrelation

The *Ljung-Box Q-statistics* is given by

$$Q = T(T+2)\sum_{j=1}^{p}\frac{r_j^2}{T-j},$$

where r_j is the *j-th* autocorrelation and T is the number of observations. *Q* is used to test the hypothesis that all of the autocorrelations are zero, i.e. that the residuals are white noise. Under the null hypothesis, *Q* is distributed as χ^2 with degrees of freedom equal to the number of autocorrelations, *p*.

Heteroskedasticity

Whites's Heteroskedasticity Test (*WHT*) can be used to test the heteroskedasticity of the OLS residuals. The test is the following. If we have a regression of the form

$$y_t = b_1 + b_2 X_t + b_3 Z_t,$$

the test is based on the augmented regression,

$$y_t = b_1 + b_2 X_t + b_3 Z_t + b_4 X_t^2 + b_5 Z_t^2$$

WHT is used to test the hypothesis that the coefficients of the variables in the augmented regression are all zero. The test is based on the *F*-statistic, which will have an asymptotic χ^2 distribution with degrees of freedom equal to the number of regressors in the test regression.

Normality

The *Jarque-Bera* statistic is used to test whether the residuals are normality distributed. The statistic is given by

$$\frac{T-k}{6}\left[S^2+\frac{1}{4}(K-3)^2\right],$$

where T is number of observations, k is the number of regressors, S is skewness and K is kutosis. Under the null hypothesis of normality, the *Jarque-Bera* statistic is distributed χ^2 with 2 degrees of freedom.

Orthogonality

In order to check whether the potential endogeneity of the right-hand side variables of the regression is a problem, the *Hausman specification test* (*H*) can be used. A Hausman test compares two sets of estimates, e.g. the ordinary least squares (OLS) and instrumental variables (IV), of the same parameters using the same data. More formally,

$$H=(\beta_{OLS}-\beta_{IV})'(Var(\beta_{OLS})-Var(\beta_{IV})^{-1}(\beta_{OLS}-\beta_{IV}).$$

The statistic H is distributed asymptotically as χ^2 with degrees of freedom equal to the number of parameters being tested. If the computed H is smaller than the critical value from the χ^2 -Table, the null hypothesis cannot be rejected.

Testing parameter restrictions

The *Wald* test is used to test which of the functional forms is preferred. For linear restrictions, the *F*-statistic and a χ^2 -statistic with associated probability (*p*) values may be used. If the restrictions are valid, there should be little difference in the fits obtained with the unrestricted and restricted regressions. Thus the calculated *F*-statistic is likely to be small, the *p*-value large, and the restrictions not rejected.

Functional form specification

The Ramsey RESET test is used to test the null of correct specification of the original model against the alternative that the powers of the fitted values have been omitted. Thus, if the original specification is

$$y = \beta_1 + \beta_2 X ,$$

the augmented specification is

$$y = \beta_1 + \beta_2 X + \beta_n Z ,$$

where Z is the linear powers of the fitted values of the dependent variables, i.e.,

$$Z = \begin{bmatrix} \hat{y}^2 & \hat{y}^3 & \cdots & \hat{y}^n \end{bmatrix}.$$

The RESET test is applicable only to OLS estimations. The test uses *F* and *LR* statistics for testing the hypothesis that the coefficients on the forecast vectors are all zero.

References

Davidson, R. and J. M. MacKinnon (1993). *Estimation and Inference in Econometrics*, Oxfrod University Press.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*, Econometric Society Monographs, Cambridge University Press.

ISBN 951-40-1525-8
ISSN 0358-4283