ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B 00120 Helsinki 12 Finland
Sarja A 11 Series

Timo Summa

INTRA-INDUSTRIAL TECHNICAL PROGRESS AND STRUCTURAL CHANGE

An application of the frontier and short-run industry production functions based on micro data

Helsinki 1986

ISBN 951-9206-11-6

ISSN 0356-7435

PREFACE

This study has grown out of a profound interest in enterprises and the regeneration of industries. The entire research process pivoted on the decision to start. Without the initial nudge from Professor Arvi Leponiemi and his unceasing and motivating support, this study would not have been begun nor been finished in such a short time. At that critical point, my decision was also influenced by discussions with Dr. Sirkka Hämäläinen, Mr. Matti Korhonen and Professor Uolevi Lehtinen. My warmest thanks to them all for their encouragement and support.

The inspiration arising from my daily working environment in the dynamic and active attitude towards the exploitation of research results at the Federation of Finnish Metal and Engineering Industries has been of vital importance and value. I am especially grateful to my superior, Mr. Harri Malmberg, for his understanding, support and willingness to accept the inconvenience caused by my six-month leave of absence during the most intensive phase of the study.

It has been a privilege to have been guided in my work by two prominent scholars in the field, Professor Finn Førsund of the University of Olso and Professor Lennart Hjalmarsson of the University of Gothenburg.

Docent Timo Airaksinen has contributed ungrudgingly of his time and advice, showing interest in both my theoretical and empirical problems. I thank him for placing his extensive knowledge and experience at my disposal. Mr. Jussi Karko has deepened my understanding of production theory and untiringly commented the numerous versions of this report.

Professor Aarni Nyberg, Docent Pekka Lehtonen and Professor Jouko Ylä-Liedenpohja have helped me greatly with numerous comments and discussions during the study process.

Part of my leave of absence was spent at the University of Southern California. Discussions there with Professor Dennis Aigner and Subal Kumbhahar, Ph.D. shed new light on especially the latest estimation methods.

The report went through several drafts involving extensive responsibility for the translator Ella Haapasalo, M.Sc. Econ., and the typist Mrs. Tuula Ratapalo, who worked with speed and skill. Mrs. Arja Selvinen has professionally displayed her talents in drawing the figures.

I wish to express my deep gratitude for the understanding shown by my parents as well as my wife Marita and my children Akseli, Johanna and Antti.

Finally I gratefully acknowledge the financial support of the Yrjö Jahnssonin säätiö, the Academy of Finland and the Osuuspankkijärjestön Tutkimussäätiö. I also thank the Research Institute of the Finnish Economy ETLA for providing me with working facilities during the reporting phase and publishing the report in its series.

Helsinki, March 1986

Timo Summa

CONTENTS

1. INTRODUCTION

1.1. Preliminaries

Efficiency is the key to success in a firm, an industrial sector and an economy. Efficiency and the related concepts technical progress and structural change are themes of interest for both the management of firms, economic organisations, financial institutions and economic and industrial policy makers. Today the economical utilisation of resources is emphasised not only by the private but by the public sector as well. There are a multitude of viewpoints and theories regarding efficiency and the related concepts and numerous tools for analysing them.

The end-use of the analysis is the decisive factor in determining, in each individual case, the relevant aggregate level. This can, for the purposes of efficiency measurement, be subdivided e.g. into the following three categories (Johansen 1972):

Efficiency at the <u>micro level</u> concentrates on the utilisation of resources of a firm or a part of it, e.g. a plant. The best practice unit serves as a basis in performance measurement. In business economics and engineering sciences, efficiency measurement on the micro level and the reasons behind the efficiency differences between individual firms have been the object of a rapidly growing interest and attention. Obviously, the particular objectives of a firm must be specified when determining whether it is efficient or not, i.e. a firm can be perfectly efficient with respect to its own objectives, but inefficient with respect to other objectives judged superior by the observer.

At the macro level, the analysis concentrates on measuring the allocative efficiency. The economic performance of an observed allocation of resources to various sectors is compared with the result of some ideal allocation, usually a Pareto-optimal one, given the existing income distribution. Another standard of reference is an allocation maximising some welfare function. Studies measuring the losses due to monopoly are examples of macro level analysis.

Measurements at the semimacro or industrial level are based on the sets of production possibilities given to each firm. The main purpose is to measure the relative performances of the firms within an industry. Each firm is, as a rule, compared with the best practice one or with a frontier production function. Efficiency differences indicate the potential output increase of the industry achievable by employing resources in firms using the best practice technology.

Traditionally industrial studies have concentrated on the "average" production unit only. In the real world, it often is not enough for an executive to compare the performance of his plant with an "average" unit within the industry. The piece of information most valuable to him, e.g. while making an investment decision, is the efficiency of the best practice unit and the factors contributing to it. Likewise it is important, while using and developing the instruments of economic and industrial policy, to know the intra-industrial efficiency distributions of the various sectors of industry.

In the neoclassical world, adaption, substitution and the determination of the optimum capacity of an industry sector are assumed to be smooth and continuous. This viewpoint is, besides being often unrealistic, awkward as a basis for structural analysis. The viewpoint suitable for our purposes would lay the main emphasis on the firm's capital, the technology linked with it, and efficiency. Capital is, as a rule, the most fixed factor and the least easily - if at all - divisible one. Change in the capital stock, i.e. investment, is bound with important strategic choices, both on the level of a single firm and an industry sector. It takes both money and time to change the production machinery. All these factors are elements of everyday decision-making.

Firms, especially in a process industry, live in a world in which technological progress can be divided into two main categories: continuous (although not even) disembodied progress and largely capital embodied one. The latter usually advances stepwise; it is linked with new vintages of capital goods and realised in the firm's major investments. The disembodied progress, on the other hand, has more to do with experience, i.e. learning curves. The classical example is the Horndal iron and steel works. According to Åselius (1957), output per working hour rose by 2 per cent per annum in 1939-1950, despite the firm's lack of reinvestment in plant and equipment. An even more pronounced case of the "Horndal effect" is presented by another Swedish steel mill: Between 1953 and 1977, the tonnage rolled per man-hour at the Hagfors strip mill rose at an average annual rate of 3.7 per cent without any net investment in plant and equipment.[1)]

Both disembodied and embodied factors are taken into account in the model developed by Johansen (1959) and Salter (1960), called a putty-clay one. In a putty-clay model, machinery and equipment can be de-

---

1) See Vinell (1981). For further examples see e.g. Helper and Lazonik (1984) and Boston Consulting Group (1974).

signed in an infinite number of technological and capacity variations at the planning stage. But once a certain technology has been chosen and the capital stock built up, the substitution of the various inputs cannot be influenced any more: the firm is a "prisoner" of its choices throughout the lifespan of the investment. It was Phelps (1963) who first gave this ex ante factor substitution and ex post nonsubstitution the name "putty-clay".

The structure of the capital services, describing the overall capacity of the industry, and the input coefficients of each individual capital item are expressed in a concise form in the putty-clay model. Investments made at various points of time are assumed to utilise the latest, technically most advanced and efficient technology of that time so that these investments represent different vintages.

The other "pole" in the dynamics of an industrial sector lies, in the Johansen-Salter framework, in obsolescence, the withdrawal from operation of firms and capital goods. A capital goods item is withdrawn from the overall capacity of the sector, i.e. becomes obsolescent, when it no more brings revenue enough to cover the variable production costs. This quasi-rent may become negative long before the capital goods item would have to be replaced because it is technically worn out. Since the latest technology utilises the input factors more efficiently and all plants, from the newest to the oldest, generally have to pay the same prices for their input factors, "economic obsolescence" decisions are part of the daily work of the corporate management and industrial policy makers.[1)]

---

1) The problems are discussed e.g. in Ballance and Sinclair (1983), with automobile and steel industries as examples.

When we give up the neoclassical representative firm theory and other assumptions of smooth and continuous adaptation in favour of the more realistic putty-clay framework, we gain a basis for defining the efficiency structure in several ways which together - complementing each other - constitute a usable analysing tool-kit. The efficiency structure of an industry is a composite of the individual performances of the micro-units. These can be compared with each other by compiling "ranking lists", based on various criteria, or by other aspects of performance, e.g. defining the best-practice plant or the production frontier.

1.2. A brief historical review of the topic

It is the market system of perfect competition that is at the centre of the neoclassical theory, not the firm per se. Its main objective is to predict changes likely to happen in the supply of products and the demand for inputs when changes occur in the only external variables that the decision-making units act on, namely the market prices for outputs and inputs. The neoclassical theory is, therefore, not a suitable tool for analysing such problems as the process of structural change in an industry in which firms differ in size and structure with regard to input coefficients and plants become obsolete when e.g. the market size increases, a non-expected factor price development takes place and embodied technical progress occurs.

Structure is a concept of no particular interest unless there is certain stability - inertia or clayishness - in the capital structure of the industry. Without inertia or immobility or non-malleability of fixed factors, no structural problem arises. It is interesting to note that there are comments on vintage aspects of industrial structure already in texts by Marx, Schumpeter and Marshall. Marx shows a great interest in the structural development of different industries, based on a genuine knowledge of their circumstances with regard to size, structure, obsolescence, labour productivity and technical progress.[1)] Schumpeter emphasises an evolutionary process based on innovations, which means rapid obsolescence and consequent destruction of any industrial structure that exists at any moment.[2)] In his theory of production Marshall introduces the concepts of quasi-rent and obsolescence; he does not, however, further develop these but turns to the analysis based on the idea of the representative firm.

Scandinavian economists have a long tradition of studying the problems of industrial structure. In 1918 Heckscher, in a report on industrial problems in Sweden, introduced a diagram in which the firm's current average costs were ranked in increasing order. On the basis of this diagram, Heckscher analysed the impact of tariff changes on industrial structure. In 1931 Åkerman studied the differences between the best-practice and average productivity of labour at Swedish saw mills. The distance between the best-practice and average-practice is also discussed in an article by Svennilson (1944).

In the light of the present study, an article by Mitchell is very interesting. He introduces the term "best current practice" and discusses

---

1) Marx's vintage theory is discussed in more detail in Hjalmarsson (1975).

2) See Elliot (1980).

the potential increase in output if all existing equipment could be transformed into best practice equipment.[1)]

The main ideas of the putty-clay production theory were proposed by Johansen (1959) and closely related ones are found in Salter (1960), with a distinction between the best-practice and average-practice productivity. A cornerstone in the development of the putty-clay production theory is Johansen's "Production Functions" (1972), in which he develops a dynamic theory of production through the integration of micro and macro and of short- and long-run aspects. This framework provides us with a chance of a deeper empirical insight into the structural change of an industry, more relevant than e.g. that obtained by an analysis based on the traditionally estimated average production function.

In the putty-clay model of Salter and Johansen there are as many different kinds of capital goods as there are time periods. A unit of a capital good of a given vintage will provide a certain capacity to produce output and will require a fixed amount of current inputs per unit of output (input coefficients). These characteristics remain unchanged throughout the life of the capital good. Technical progress then implies that capacity of a later vintage will always be more efficient than that of an older vintage.

Førsund and Hjalmarsson have deepened the Johansen-Salter framework both theoretically and empirically in the 1970s and 1980s. They have emphasised the dynamics of the concepts of structure and structural

---

1) See McKenzie (1937), p. 119.

change[1] and the role of technical progress and introduced new, more general measures for efficiency measurement and more developed estimation methods. They have brought the Johansen-Salter approach into the industry studies as a useful analysing tool.

Besides the theoretical framework in economics, new estimation methods have been developed, especially stochastic frontier models. A brief overview of this development will be found in Section 2.3.2.[2] In addition, many modern studies in the area of productivity are based on the theory of economic index numbers connected with flexible functional forms of production functions, pioneered by Diewert in the mid-1970s.[3]

1.3. Main objectives of the present study

We will study efficiency on the micro level and its linkage with the industrial efficiency structure within the Johansen-Salter framework, enabling us to include the non-representative firm an apply a clearly defined concept of an industry in our analysis. We advocate the view that the neoclassical notion of a representative firm with smooth substitution possibilities has to be replaced by a putty-clay framework yielding a more realistic modelling of actual, intra-industrial development patterns.

---

1) See Førsund and Hjalmarsson (1974).

2) For a survey, see Schmidt (1985).

3) Fischer (1984) applies this approach to measuring total factor productivity, efficiency and technical change in the West-German industry.

We apply precise definitions of the efficiency concepts (output and input saving) and present extensions of the theory of efficiency measurement by applying new industry level efficiency measures linked to the short-run production function.

The main contribution is that we for the first time apply the two production function concepts, the frontier function and the short-run industry production function (SRIPF) simultaneously, on the same data set. In addition to the deterministic approach, we also look at the stochastic one and estimate several frontier models based on these approaches, in order to study their relevance in different situations. New measures for decomposing the productivity growth at the industry level are applied.

We empirically test different specifications of the frontier production functions and the SRIPF and construct the efficiency measures based on these in different development phases of the Finnish brewing industry. We have collected and worked upon high quality data and a long time series needed in a demanding analysis. The time series includes development phases with markedly differing characteristics. For the purposes of this study, alternatives were constructed for operative variables, enabling us to analyse the sensitivity of the empirical results to the various specifications of the variables.

We construct a many-sided and concrete picture of the efficiency structure and technological progress of the observed industry during the various, strongly contrasting phases of the observation period.

1.4. Design of the report

Chapter 2 constitutes the theoretical base of the study. First we introduce the development of the putty-clay theory of production with special emphasis on Johansen's integrated production function system of ex ante and ex post functions. We stress the point that, within the framework chosen, the connection between short-run industry functions over time goes through the ex ante functions of the micro units. The presentation of different approaches arising from the theory of production will be sufficiently general to accommodate any specification of the production, cost and profit functions that satisfy the usual regularity conditions required by the duality theory. Then several efficiency concepts useful in an integrated production function system are reviewed. We finish the theoretical part by discussing the means of estimating the ex ante function via estimating frontier production functions, and by establishing short-run industry production functions. Regarding the former ones, various programming methods and maximum likelihood for composed error stochastic frontier models are discussed. As regards the latter, Johansen's approach in the discrete case for establishing short-run industry production functions is outlined together with further development of the means of characterising such functions, based on Førsund and Hjalmarsson.

Chapter 3 contains the empricial application of the study. The models are applied to data for the Finnish brewing industry over the period 1955-1984. We use both deterministic and stochastic specifications in the calculations and estimations of the frontier production function. Several measures for technical and cost efficiency are presented. The properties of the short-run industry function are illuminated in different ways.

In Chapter 4 we discuss the measures obtained for structural efficiency on both types of production functions. The computed short-run cost functions are utilised within an index framework to identify the contributions from pure technical progress and from factor price increases.

In Chapter 5 we present some concluding remarks as well as some suggestions for further research.

In the Appendices we present i.a. a description of the data used and the working of the algorithm for constructing the short-run industry production function.

2. PRODUCTION THEORY AND THE ANALYSIS OF TECHNICAL PROGRESS AND STRUCTURAL CHANGE

2.1. Analytical framework: An integrated system of production functions

This study concentrates on the measurement of short- and long-run intra-industrial efficiency on the basis of micro data, by using production functions. Starting from micro units is relevant, because in the real world there is no single industry (or macro) decision-maker who attempts to maximise profits or minimise costs to allocate resources optimally on the basis of the industry production function (except in the case of pure monopoly). In our case the links between the micro and industry levels are highly relevant, since we study micro foundations of macro relations. Another important relation to be treated is that between ex ante and ex post and between the long-run and the short-run decision making.

In this chapter we point out the shortcomings of the neoclassical production theory for empirical analysis. As an alternative we introduce the putty-clay approach and the integrated production theories by Johansen and Sato based on it. Finally, we shortly discuss the joint potential of Johansen's ex ante function at the micro level and the short-run production function at the industry[1] level.

---

1) We use "industry" instead of Johansen's notion "macro", but the meanings are identical.

The neoclassical theory of the firm was originally intended to connect the firms to the market and, with its assumptions of smooth, costless possibilities of substitution and choice of optimal scale, it fits well to the analysis of stationary states and to the long-run development of industrial structures at an aggregated level. It is, however, not a theory suitable for short or medium term analysis of intra-industrial structure. Attempts have, however, been made to apply the neoclassical theory on the analysis of industrial structure, but without major success.

The traditional empirical approach to production theory is based on the assumption of a representative firm, which as a theoretical concept was launched by Marshall starting in the 1890's and Pigou after the First World War.[1)] The firms are supposed to share an identical production function. In econometric applications, based on data from real micro units, the firms are supposed to differ in efficiency from each other in a neutral and stochastic way. This latter approach is due to Marschak and Andrews (1944). The presupposition that neutral differences in efficiency are distributed normally for one reason or another (mostly unexplained) dispenses with the necessity of introducing an aggregation principle for micro-production functions to yield a cross-section production function since all micro production functions can be reduced, after eliminating neutral differences, to a single production function. This assumption enabled Marschak and Andrews to concentrate on the statistical identification of the func-

---

1) Moss (1984) gives a compact survey of the development of the theory of the firm and the characteristics and shortcomings of the representative firm concept.

tion in a simultaneous equation system derived from the first order conditions of profit maximisation. Thus the main emphasis was shifted towards problems dealing with simultaneity among the variables and, as there is only one production function to worry about, the problem is greatly simplified.

The approach exerted a profound influence on subsequent developments of econometric methodology in this field. However, at the same time the econometricians paid little attention to the realism of the assumption itself. They forgot to ask such questions as whether micro production functions really differ in the way postulated by the assumption and to stress a methodological point: are the theoretical underpinnings of stationary disturbances valid in the circumstances. The whole test theory of the related econometrics is built around the hypothesis of normally distributed stationary disturbance around the average function both in the case of the production function and in that of dual cost or profit functions. The realism of the theory does not escape behind the new or popular flexible functional forms.

As a description of an industry, the representative firm is the closest approximation to the average firm and the empirical estimation result is basically a micro relation. Although its internal construction is assumed optimal from the cost minimisation point of view, its inherent, totally homogeneous efficiency, such a "featureless" construction description is not very useful as a tool for analysing the structure of an industry and not suitable, for instance, for measuring differences in efficiency within an industry, since differences in efficiency are stochastic by definition.

However, empirical observations tell us that not all firms within an industry operate with the same degree of technical and economic efficiency. Moreover, the differences between micro units are, in practice, not random but, on the contrary, results of conscious actions and autonomous evolution. Therefore, one might ask whether the application of the representative-firm theory in practice is justified.[1)]

It is apparent that the neoclassical theory has, to a large degree, exhausted its potential in analysing the industrial structure and the changes occuring in it. Emphasis on realism and the practicability associated with it have created a need for developing a production theory recognising the fact that firms do differ in productive efficiency as a starting point.

Heckscher (1918) emphasised the differences between firms in an industry by using a distribution that presents the micro units ranked according to the size of their variable unit costs, calculated on the basis of observed prices. This distribution describes the economic features of efficiency in terms of an economic ranking. Salter (1960) used the same principle to construct his famous diagrams ranking firms by their individual input coefficients along the ordinate axis and the accumulated relative capacity of the firms along the abscissa axis. These diagrams describe partially the physical or technical aspects of the efficiency for one input. Contrary to Heckscher's cost approach, the Salter diagrams do not give a unified picture of physical efficiency,

---

1) Robbins (1928, p. 393) already criticised the neoclassical theory: "There is no more need for us to assume a representative firm than there is for us to assume a representative piece of land, machine or worker." Note that a representative firm is not restricted to be in perfect competition. The concept may also be formulated in a monopolistic situation as well. This was also known by Marshall and Pigou. See Moss (1984).

for depending on the input the efficiency ranking may differ from one input to another.

Salter's approach in analysing the intra-industrial structure is based on the assumption of embodied technical change in a model with ex ante factor substitutability but ex post nonsubstitutability i.e. putty-clay, according to Phelps' (1963) terminology. Embodiment means that technological progress is related to fixed capital, i.e. a prerequisite for technical change is investment. Consequently there are as many different kinds or vintages of capital goods as there are time points on the investment path. A unit of a capital good of a given vintage will provide a certain capacity to produce output and will require a fixed quantity of current inputs per unit of output. These characteristics may remain unchanged throughout the life of the capital good. An idealised putty-clay approach with embodied technical progress then implies that the capacity of a later vintage will always be more efficient than that of an older vintage. Thus the putty-clay model differs essentially from the neoclassical theory, according to which there are no problems in changing the input mix even when the investment is made. In other words, while capital is treated as homogeneous in the neoclassical model, the putty-clay model recognises the heterogeneity of capital, a fact of the real world. Another major advantage of the putty-clay model is that it brings obsolescence of capital into the analysis, a feature excluded from the neoclassical model.[1)]

A relevant production theory, based on the putty-clay approach, on the other hand, presupposes a successful solution to the aggregation problem. In some cases such a solution can be found.

---

1) See Førsund and Hjalmarsson (1984).

Houthakker (1955) was the first to tackle the aggregation problem with the distribution of fixed input coefficients and capacities as his starting point. He derived a Cobb-Douglas industry production function on the basis of a generalised Pareto distribution of production units with respect to input coefficients and capacities instead of starting from a collection of identifiable individual production units.[1)]

Johansen (1972) elaborated further Houthakker's special case and constructed a production theory[2)] based on this "distribution approach", the putty-clay framework. A very rigorous mathematical formulation of this production theory is provided by Hildenbrand (1981) and Seierstad (1985). An additional contribution is made by Sato (1975). A common feature in the theories of Johansen and Sato is that the production functions can be divided into micro and macro (= industry) level functions and the efficiency distribution can be used for deriving mathematical relations between these. Both Johansen and Sato emphasise the efficiency distribution's function as the strategic link between micro and macro economic behaviour in any realistic circumstances.

Johansen's (1972) integrated production theory consists of four different concepts of production functions:

1 The ex ante function at the micro level
2 The ex post function at the micro level
3 The short-run function at the industry level
4 The long-run function at the industry level

---

1) See Houthakker (1955) and Hildenbrand (1981).

2) The empirical studies by Heckscher as well as Salter were not performed in a formal production theory framework.

At the micro level the ex ante function is a kind of planning function including all the potential technologies at hand at a certain point in time.[1] The ex-ante function is in fact the efficient envelope of all the alternative technologies conceived of by designers and available to managers at the time when the investment decision is made. It is assumed that at this stage there is a free choice of capacity and substitution possibilities between all inputs, depending on the producers' "planning prices", future expectations and other planning aspects. The ex post function at the micro level sums up the choices made by the firms from the available ex ante functions, determining their layout and their adjustment to external changes. In general it is supposed that after the technology chosen from the ex-ante possibilities has been installed, the capacity is given and there is no room for substitution; the ex post function is, therefore, characterised by fixed production coefficients.

Even at the macro level, two separate cases can be found in Johansen's production function framework. The short-run industry production function (SRIPF) is defined by maximising the industry output for given amounts of industry inputs subject to micro technologies and fixed capacities of the micro units as expressed by the ex post production functions. The SRIPF at the macro level will, at any given moment of time, comprise a certain number of production units and it aggregates the ex post functions of the micro units into a short-run production function according to the efficiency criterium. This means maximising of the output of the industry with given micro-units and it can be shown that with given prices this criterium corresponds to cost

1) In the words of Salter (1960), p. 15 "... the production function which includes all possible designs."

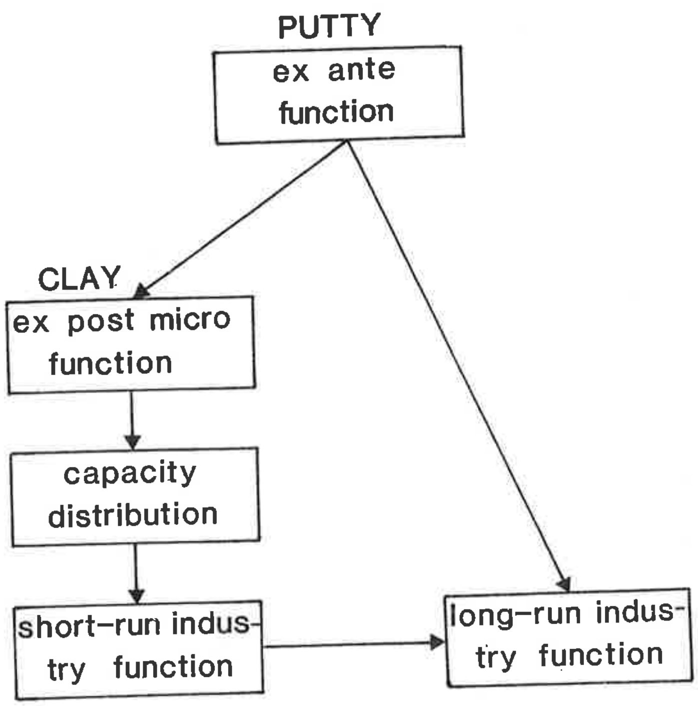
minimisation for given output levels. It is assumed that each of the production units has different input coefficients and different production capacities. Total production at industry level will be obtained as a result of an efficient combination of activities given by the existing ex post micro production functions. The short-run production function yields the optimal way of utilising all production units of an industry as a function of the degree of capacity utilisation and relative factor prices.

Another macro level function is the long-run industry production function, which is closely connected with the ex ante function at the micro level. Johansen regards the long-run production function as a more hypothetical construct. It is assumed to correspond to that perfectly efficient production technology, which fully utilises available supplies of inputs of the industry. In an industry where the ex-ante technology is continuously improving and factor supplies changing over time, the macro function is unlikely to be static. The function is constructed on the basis of factor supplies available to the sector at a particular point in time. In contrast to the short-run industry production function, we here make the hypothetical assumption that capital is malleable and can take on any desired form. Under these conditions we maximise output as a function of capital and current inputs. This function is the long-run production function for the sector.

How does the long-run production function relate to the short run one? Only in the case where the short-run production function has all capacity concentrated in a single point in the input coefficient space will a point on the long-run production function be realised. As soon as we have some scattering in the capacity distribution, the points that are realised from the short-run industry production function will become inferior to the points on the long-run function, provided that the ex ante function is not linear.

The actual production potential of the industry as a whole at a given moment of time is, however, not determined by the long-run production function, but rather by the short-run production function at industry level, which is dependent upon the technical characteristics and capacities of existing production units.[1] Bosworth (1976) describes Johansen's scheme and the putty-clay concept in it by using Figure 2.1.

Figure 2.1: Johansen's integrated production function system



---

1) Johansen shows a clear mathematical relationship between the ex ante production function at the micro level and the long-run production function at the industry level at any given moment of time. If the ex ante function is a homothetic function then the homogeneous kernel of the ex ante function appears as the long-run industry function. See Johansen (1972), p. 6-25.

Sato's production function system is rather similar to that of Johansen. Sato makes a clear distinction between alternative concepts of the production functions and uses the following seven variants:[1)]

1 ex ante production function

2 micro production function in the short-run

3 micro production function in the long-run

4 macro production function in the short-run

5 macro production function in the long-run

6 macro production function at full-capacity

7 cross-section production function, frontier

In terms of the elasticity of substitution in a two factor case (σ), Sato's production functions are ordered as follows:

$$\sigma_5 \geq \sigma_1 = \sigma_3 = \sigma_6 \geq \sigma_7 > \sigma_4 > \sigma_2 .$$

Sato was influenced by the exact aggregation conditions of Fisher (1968) and (1969) and tried to avoid these difficulties by including a theory of approximate aggregation of micro units in a short-run production function at industry level. The main point of difference between the theories of Johansen and Sato lies in the fact that Sato handles capital as an explicit variable, whereas in Johansen's approach it is supposed that capital (services) is "compatible" with capacity. For this Sato needs a more sophisticated system of production functions, but basically their main principles of connecting the micro and macro levels remain the same. If we assume a putty-clay-technology the number of possible production functions are reduced to Johansen's four concepts.

---

1) See Sato (1975), p. 141.

In the following we only use the Johansen concepts which are clearly defined and fully integrated, compared to those of Sato. Two functions in the Johansen framework are especially relevant and serve as the theoretical basis of this study:

1 The ex ante function at the micro level and
2 The short-run production function at the industry level.

The interpretation of the ex ante concept is not obvious and one can think of different types of ex ante functions. Our concept of an ex ante function is based on actually observed best-practice technology expressed by the so called frontier production function. In the vintage sense it is close to the Johansen ex ante micro level function which is associated with engineering know-how and provides a norm against which the efficiency of an individual firm (micro unit) can be measured. Even structural efficiency measures can be constructed analogously for the entire industry (See Chapter 4).

For a given set of input prices the short-run industry production function yields the order of utilisation of the micro units along the expansion path and thus the marginal cost curve. For the same set of prices the Heckscher diagram yields the same ranking of the units identical to the order in which the units appear in the marginal cost function. The Johansen short-run production function at macro level solves the problem of the efficient utilisation of the production units in industry when the rate of capacity utilisation increases from zero to 100 per cent in the industry.

The connection between a series of short-run industry production functions over time goes through the ex ante production functions of the micro units with the fixed factors as variables. The ex ante function is the choice of technique function for the construction of an individual micro unit. The short-run industry production function reflects both the history of ex ante functions over time and the actual choices made from these ex ante functions. Production at any point of time must be compatible with the short-run function.

The ex ante production function at micro level and the short-run production function at industry level are parts of the same system and complement each other well as tools for describing intra-industrial technical progress and structural change.

In Sections 2.2. and 2.3. we will study the ex ante production function at micro level defined as a frontier production function, which will serve as a basis for measuring industrial efficiency when all factors are variable. Correspondingly, in Section 2.4. we apply the short-run production function at industry level when only current inputs are variable so that fixed factors, such as capital, only determine the capacity of the individual micro units.

2.2. Different approaches arising from the theory of production

It is only in relation to specific objectives that the measurement of economic performance can be meaningful. Johansen's (1972) integrated theory of production is based on the thesis that the different production function concepts are relevant to different levels of aggregation and over different periods of time. We have pointed out that two of Johansen's four production function concepts serve as a relevant theoretical basis in the analysis of intra-industrial technical progress and structural change. First we concentrate on the measurement of efficiency and technical change using the frontier function framework, which simulates Johansen's ex ante function at micro level. In this section we give a brief discussion of the definition of the frontier function based on the production function and show that we can define cost and profit frontiers in rather similar ways.

We need a standard against which to measure efficiency. To say that a firm or a plant produces x per cent of its potential output, given its input usage, we need to know what the maximal 100 per cent output is. This reference base may be derived from the firms' pursuit of optimal performance, i.e. minimisation or maximisation of the objective function. In this study we will use a frontier production function with the word "frontier" emphasising the idea of maximality which it embodies. The theoretical notion of Johansen's ex ante micro function, as referred to earlier, is that it represents the most efficient means of transforming inputs into outputs.[1] Ex ante functions based on observed performance are usually called frontier production functions while those based on engineering knowledge are called engineering production functions although the ex ante and frontier concepts often are regarded as synonymous.[2] In this study we deal with frontier functions based on observed performances.

In the literature there are three alternative ways to derive the frontier functions. The production function shows the maximum output ob-

---

1) Johansen's ex ante function-concept corresponds to the blueprint technology in Grosse (1953) and the best-practice technology in Salter (1960).

2) See Førsund and Hjalmarsson (1984), p. 114.

tainable from given input vectors. The cost function gives the minimum level of cost at which it is possible to produce some level of output, given input prices. The third approach, linked with the above-mentioned ones, is that of profit maximisation. In the latter case there also is a profit function, which gives the maximum profit attair 'le, given output price and input prices. The linkages between the three functions are discussed below on the basis of the compact surveys of Førsund, Lovell and Schmidt (1980) and Diewert (1974). The frontier concept can be applied to all these three functions. In the literature they are developed within the neoclassical framework. We will first give an exposition of the three functions and then comment on their relevance to the putty-clay world.

When introducing the frontier production function we assume, for simplicity, that all m firms of the industry studied produce a single homogeneous output y, from a vector of inputs, x, consisting of current inputs and capital. The production function f(x) summarises the technology available to the firm. Technology conveys the production possibilities which are open to the firm when transforming productive inputs $x \equiv (x_1,...,x_n)$, available at fixed prices $w \equiv (w_1,...,w_n) > 0$, to a single output y that can be sold at a fixed price $p > 0$.

The production function $y = f(x)$ describes the maximum output obtainable from a given input vector.

It is defined for non-negative output and inputs and satisfies usually certain regularity conditions[1]:

---

1) Cf. Diewert (1974).

(i) A positive amount of product can be produced from a ector of non-negative amounts of inputs.

(ii) f is a continuous, twice differentiable, non-decreasing function of inputs. (1)

(iii) f is a quasiconcave function, that is technology, summarised by the production function, represents non-increasing marginal rates of substitution.

If the production function fulfills these conditions we may also define the producer's total minimum cost function c(y;w) as the solution to the following constrained minimisation problem:

$$c(y;w) = \min_x [w'x:f(x) \geq y] \tag{2}$$

in which w' is the transpose of the input price vector.

In other words, the producer takes prices as given and attempts to minimise the cost of producing a specified output level, y. Total cost thus depends on the chosen output level y, the given vector of input prices w, and the given production function f. The cost function has same kind of regularity properties with its arguments as the production function: it is positive, nondecreasing, homogeneous of degree one and concave, continuous function in w for every fixed y. Moreover, it is continuous and non-decreasing in y for every fixed w.[1)]

---

1) Thus the condition of quasiconcavity of the production function in x is replaced by a stronger condition of linear homogeneity and concavity with respect to input price vector w; cf. Varian (1984) p. 45, Diewert (1982), p. 541.

The production function f thus determines the cost function c through definition (2). The converse is also true: a cost function satisfying regularity conditions given above determines a production function:

$$f(x) \equiv \max_y \, [y : c(y;w) \leq w'x \text{ for every } w] \qquad (3)$$

where $x \equiv (x_1, x_2, \ldots, x_n)$ is a given vector of inputs and c is the given cost function so there is duality between cost and production functions. Given one of these functions, under certain regularity conditions, the other can be uniquely determined, a result originally due to Shephard and Samuelson.[1)]

This is a nice theoretical result, but it is the following result, usually called Shephard's lemma which makes the duality theory extremely useful in empirical applications[2)]:

$$\frac{\partial c(y,w)}{\partial w_i} = x_i(y;w) \text{ for all } i. \qquad (4)$$

The value of function $x_i(y;w)$ is the cost minimising quantity of input i needed to produce y units of output given input prices w. These functions, called conditional factor demand functions, are homogeneous functions of degree zero with respect to input prices and monotone in output. Changes in cost minimising input demands induced by an increase in output can not all be negative, i.e. not all inputs can be inferior. This follows directly from the non-decreasing-in-y-property of the cost function via Shephards lemma.[3)]

---

1) See e.g. Samuelson (1947), Shephard (1953) and Diewert (1974).

2) See e.g. Varian (1978).

3) See Diewert (1982), p. 568.

Shephard (1953) also introduced the concept of a homothetic production function and determined the properties of the corresponding cost function. Homothetic production functions are defined as being of the form $F[f(x)]$, where f satisfies conditions (1) but is also homogeneous of degree one with respect to inputs.[1)] F is a monotonically increasing, continuous non-negative function. In this case, Shephard showed that the corresponding cost function can be separated into the form $c(y;w) = F^{-1} [y] \bar{c}(w)$, where F^{-1} is the inverse function of F and $c(w)$ the price function, which is homogeneous of degree one with respect to input prices (irrespective of the homogeneity of production function).[2)]

We have two distinct methods of deriving a system of cost minimising factor demand functions: One would be to postulate a well-behaved functional form for a production function and then use mathematical programming techniques in order to solve the cost minimisation problem to obtain the factor demands from the first order conditions for the cost minimum.[3)] Inserting the factor demands into the cost equation yields a cost function with input prices and output as arguments. The second method would be to postulate a well-behaved functional form for a cost function and using duality to obtain the system of conditional factor demand functions directly via Shephard's lemma by partial differentiation.

1) Note that if f is homogeneous of any degree with respect to inputs, it is always possible to transform this kind of function to linear homogeneity.

2) In index-theory this is often called also unit-cost function; see e.g. Diewert (1974), (1981) and Samuelsson and Swamy (1974).

3) For a detailed exposition see e.g. Lehtonen (1976). Chapter 2.

The difficulty with these methods is that in the first approach it is usually impossible to obtain the conditional factor demand functions as explicit functions in the parameters of f. In the second method it is generally impossible to obtain an explicit form of the production function, except in special cases.

So far we have assumed cost minimising behaviour for the firm. Next we tackle the problem of obtaining functional forms for factor demands and supply functions in the context of a profit maximising firm. Given the fixed prices of inputs and output, we assume that the producer chooses the input-output combination which maximises his profit: it is well known that the firm's production function determines its profit function Π. Furthermore, McFadden (1978) and Gorman (1968) have shown that if f satisfies certain regularity conditions[1)] then the profit function Π may be used to determine the production function; moreover, a counterpart to Shephard's lemma holds for the profit function, and factor input demand and output supply functions can be obtained simply by differentiating the profit function with respect to prices (Hotelling's lemma[2)]).

For econometric applications, it is convenient to introduce the concept of the short-run profit function assuming short-run profit maximisation. The variable profit function gives us the maximum profits the firm can obtain, allowing a subset of inputs x and output y to be

---

1) In the case of profit maximisation the condition of quasiconcavity should be replaced by concavity.

2) See Varian (1978).

variable (we will refer to these quantities as variable inputs) while another subset of inputs v is held fixed (fixed inputs).[1)]

Let us now suppose that the producer's fixed inputs are fixed at amounts represented by a non-negative vector $v' = (v_1,...,v_m)$ and that he can buy or sell variable inputs or/and output at the given positive prices $(p,w)' = (p,w_1,...,w_n)$.Then the producer's variable profit function may be defined as follows

$$\Pi(p,w,v) \equiv \max\ [py - w'x | f(x,v) \geq y,\ v \geq 0] \qquad (5)$$

Thus the variable profit function depends not only on the vector of variable input prices w and output price p, but also on the vector of fixed inputs v. On the other hand, if we are given a variable profit function satisfying certain regularity conditions, then we may define the production function which corresponds to Π as follows:

$$f(x) = \max_y \Pi[y: (p,w,v) \geq py - w'x] \qquad (6)$$

The variable profit function which corresponds to f via definition (6) coincides with the initially given variable profit function (5). Thus there is a duality between the production function f and the variable profit function Π.

In a manner analogous to Shephard's lemma, the following result, called Hotelling's lemma, may be shown:

---

1) This kind of profit function is sometimes called a restricted profit function, see e.g. articles in Fuss and McFadden (1978). Fixed inputs could be introduced in the cost functions as well.

If a variable profit function $\Pi(p,w,v)$ satisfies certain regularity conditions[1] then we have for the profit maximising output supply function u

$$\frac{\partial\Pi(p,w,v)}{\partial p} = u(p,w,v)$$

and for profit maximising factor demand functions u_i, $i = 1,\ldots,N$

$$\frac{\partial\Pi(p,w,v)}{\partial w_i} = u_i(p,w,v) \text{ for all } i = 1,\ldots,N$$

given prices (p,w) and fixed inputs v.

Hotelling's lemma may be used in order to derive systems of variable output supply and factor demand functions. We need only postulate a well-behaved functional form for the profit function $\Pi(p,w,v)$.

Thus there are two different procedures for deriving a system of profit maximising input demand and supply functions. The conventional approach is based on the method of postulating a functional form for a production function and solving factor demands from the first order conditions for profit maximum as functions of output and input prices and inserting them into the production function to obtain profit maximising supply as a function of all prices.[2] The corresponding cost function is obtained by inserting the factor demand functions into the cost identity.[3] Further, the profit function is obtained by inserting the

1) See e.g. Diewert (1974) or McFadden (1978).

2) See e.g. Intriligator (1971).

3) See e.g. Diewert (1974), Lau (1978) and McFadden (1978).

supply and factor demand functions into the profit identity. This procedure is the so-called direct method.

The profit maximising supply and demand functions can, however, also be solved by an indirect method by postulating a suitable functional form for the profit function, and applying Hotelling's lemma to obtain profit maximising factor demands and supply functions.[1)] The corresponding cost function is obtained by inserting the factor demand functions into the cost function.

In the following we assume that the set of fixed inputs v is empty. So the profit function $\Pi = \Pi(p,w)$ is a variable profit function.

To discuss the relevance of frontier concepts developed above for the putty-clay framework we have to consider a typical investment decision in the putty-clay model.[2)] The choice of factor proportions and capacity depends on the entire set of expected future prices and technological development. Current cost and profit functions cannot reveal the true ex ante technology. We may conclude that only the production function estimated on the basis of input and output data is relevant. Current prices are not enough because they are only a small part of the relevant price set. On the basis of observed current inputs and outputs it should be possible, in principle, to establish a frontier function.

---

1) See e.g. Frisch (1965).

2) See for a deeper discussion Førsund and Hjalmarsson (1984), Chapter 2.

2.3. Frontier production functions: Efficiency and technical progress

2.3.1. Farrell's concepts and efficiency measures

In this section we define some of the key efficiency concepts some of which are used in the empirical part. Firstly, we introduce the basic efficiency concepts originally developed by Farrell in his pioneering work of 1957. Next we present the generalised Farrell-measures as proposed by Førsund and Hjalmarsson (1974 and 1979).

Let us start with a brief presentation of Farrell's original efficiency concepts. Farrell (1957) defined three efficiency concepts:

1 technical efficiency

2 price or allocative efficiency

3 total or overall efficiency

Technical efficiency

The general meaning of technical efficiency can be formulated as follows. Let us suppose that a firm is being observed with an output-input vector (y^0, x^0). If f is the frontier production function such a vector is said to be technically efficient if $y^0 = f(x^0)$, and technically inefficient if $y^0 < f(x^0)$. The situation $y^0 > f(x^0)$ is assumed to be impossible. One measure of the technical efficiency is the ratio $0 \leq y^0/f(x^0) \leq 1$ Technical inefficiency causes excessive input usage, which is costly, and so $w'x^0 > c(y^0,w)$. Since cost is not minimised, profit is not maximised and so $(py^0 - w'x^0) < \Pi(p,w)$.

When defining his concepts of efficiency Farrell restricted the production function to be homogeneous of degree one, i.e. there is no increasing or decreasing returns to scale. Transformed into the input coefficient space the entire production function is thus represented by a single unit isoquant. The isoquant is called the efficiency frontier and is the border towards the origin of the set of feasible input coefficients. See Figure 2.2.

Figure 2.2: Farrell's efficiency concepts based on frontier technology characterised by the unit isoquant UU'.

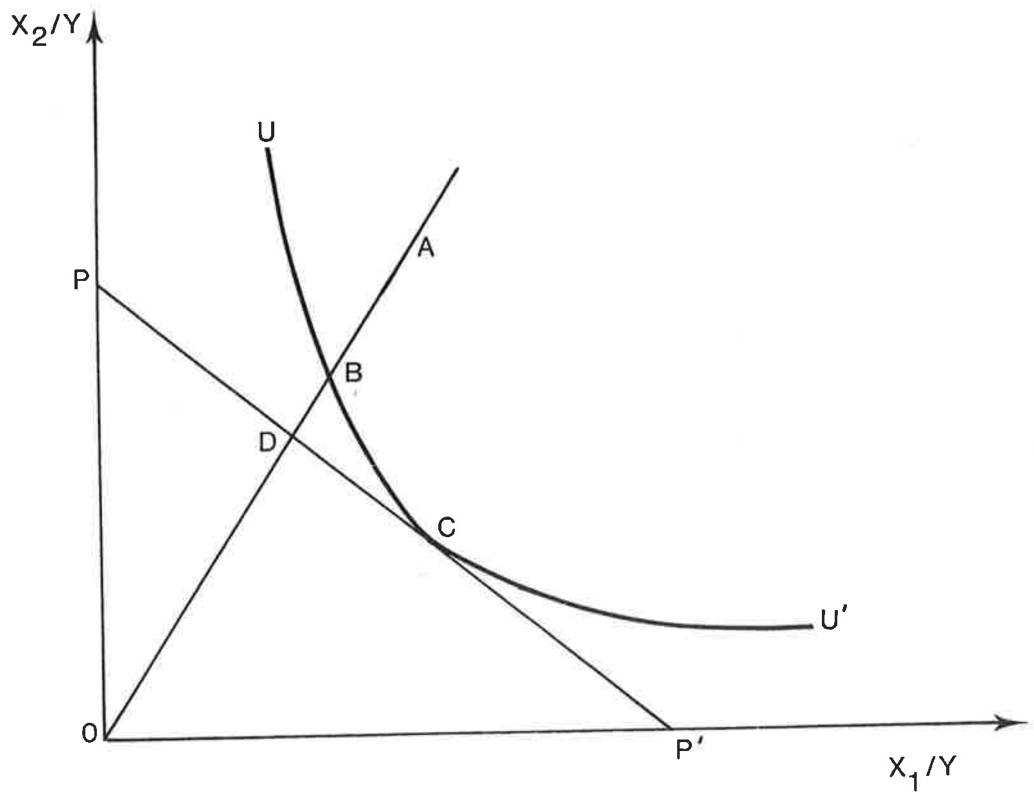


Figure 2.2 depicts an industry producing a single output y with two inputs x_1 and x_2. Under the assumption of constant returns to scale or, equivalently, linear homogeneity, the frontier production function $y = f(x_1,x_2)$ can be written as $1 = f(x_1/y, x_2/y)$. This is described by the "unit isoquant" or efficiency frontier UU'. For given input prices the point C denotes the minimum cost combination of the two

inputs, required for producing a given output. This is the point where the relative price line PP' is just tangential to the isoquant curve. If a firm in the industry is observed using (x_1^0, x_2^0) to produce y^0, let the point A represent $(x_1^0/y^0,\ x_2^0/y^0)$.The distance OA relative to OB shows how the same output could be produced with frontier technology using less inputs. The ratio OB/OA is Farrell's measure of technical efficiency and shows the ratio between the observed amount of inputs and those required with frontier-function technology, or the relative reduction in input requirements in producing the observed output by frontier production technology using the same factor proportions.

Allocative or price efficiency

In an analogous way, the ratio OD to OB in Figure 2.2 measures the deviation from optimal adjustment to factor prices, i.e. the ratio between the minimised average cost and the observed average cost. This second ratio is called allocative or price efficiency measure. The vector $(y^0,\ x^0)$ is said to be allocatively efficient if $f_i(x^0)/f_j(x^0) = w_i/w_j$ $i,j = 1,\ldots,n,\ i = j$ assuming f to be differentiable. Allocative inefficiency results from employing inputs in wrong proportions. Inefficiency is costly, and so $w'x^0 > c(y^0,w)$. Since cost is not minimised, profit is not maximised, and so $(py^0 - w'x^0) < \Pi(p,w)$. Farrell's measure of price efficiency is, however, of limited interest because the cost-minimising proportions are independent of the scale of production in the case of homothetic production functions only. Moreover, under putty-clay assumptions factor proportions should not be adjusted to today's prices but to the whole vector of expected prices in the future.[1]

---

1) See Johansen (1972).

Total or overall efficiency

Farrell used the product of the two measures for measuring the total or overall efficiency

$$\frac{OB}{OA} \times \frac{OD}{OB} = \frac{OD}{OA}$$

This ratio measures the overall economic efficiency of resource utilisation at point A in comparison with the optimum at C. It is obvious that C is a point of maximum efficiency given the assumptions where the efficiency index has a value of unity. The inherent weakness of the price efficiency measure also applies to the total efficiency measure.

With constant returns to scale a 100 per cent technical and allocative efficiency is necessary and sufficient for $(py^0 - w'x^0) = \Pi(p,w)$. If the constant returns to scale assumption does not hold, the combination is insufficient because the firm could still be scale inefficient. It follows that $(py^0 - w'x^0) = \Pi(p,w)$ if, and only if, the firm is technically, allocatively and scale efficient. If $(py^0 - w'x^0) < \Pi(p,w)$, this difference may be due to any combination of the three types of inefficiency.

Farrell originally developed his measures within the framework of linear homogeneity. For this reason there are several weaknesses in them and the same are remined also in a more general homothetic functions. This inherent weakness is basically caused by the fact that, in the case of homothetic functions, the cost function decomposes into two independent parts, as shown earlier in Section 2.2, the volume part depending only on output volume and the price part depending only on prices so that the scale function is constant or dependent on output only and the expansion paths are linear.

Førsund and Hjalmarsson (1974 and 1979) have generalised the Farrell measures to non-homogeneous and non-homothetic production functions, and non-uniform isoquant spacing. Following Farrell, Førsund and Hjalmarsson apply radial measures, i.e. the distance between an observed unit and the point of reference is measured along a factor ray. An argument in favour of applying radial measures is that such measures have a straightforward economic interpretation, can be interpreted in terms of potential input saving, potential output increasing, potential cost reduction etc. In the case of a non-homogeneous production function two measures of technical efficiency, i.e. input saving and output increasing, respectively, and three measures of scale efficiency are defined. We use Figures 2.3 and 2.4 to illustrate these concepts.

In Figure 2.3, a section of the production function is represented by the curve $y = f(\mu x^0)$ and the production unit observed at D' uses inputs x^0 and produces output y^0. Output per unit of input is maximised when a ray from the origin is tangential to $f(\mu x^0)$ as at A', where $\mu = \hat{\mu}$, output is $\hat{y}$ and the scale elasticity (ε) is unity. This is the technically optimal scale level along the chosen factor ray.[1] B' and C' are points on $f(\mu x^0)$ corresponding to a unit producing the observed output y^0 with minimum inputs $(\mu_1 x^0)$ and to one producing maximum output, y^*, with actual inputs x^0, minimum and maximum referring to frontier technology.

In Figure 2.4, the optimal scale of the production function is transformed into the input coefficient space. Point A, corresponding to A'

1) See Frisch (1965).

Figure 2.3: A section of the frontier production function $y = f(x)$ along the ray $x_1 = \mu x^o$

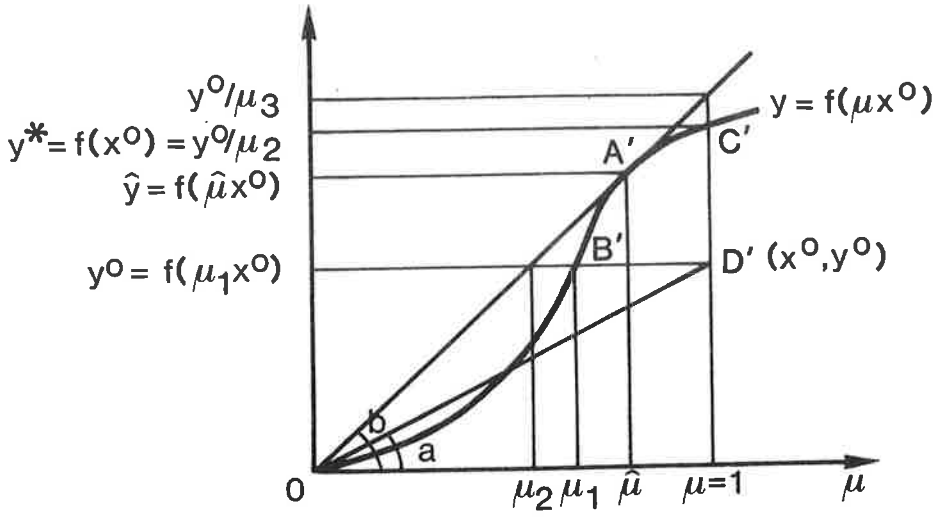
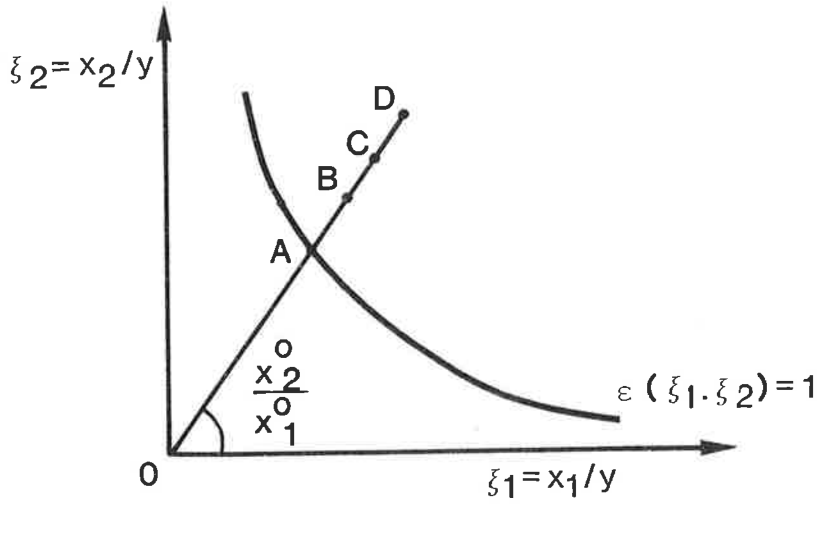


Figure 2.4: The efficiency frontier



in Figure 2.3, lies on the efficiency frontier. B and C are the transformed points B' and C' of the production surface in Figure 2.4, corresponding to output levels y^0 and y^* respectively, and D is the observed point $(x_1^0/y^0, x_2^0/y^0)$ corresponding to D'. The slope of the ray OD is x_2^0/x_1^0.

Technical efficiency

The input saving efficiency measure shows the ratio between the amount of inputs required to produce the observed output with frontier function technology and the observed amount of inputs.

According to Figure 2.3, comparing the observed point D' with the point on the frontier production function gives the input saving measure $E_1 = \mu_1$, where μ_1 is found by solving for μ_1 in $y^0 = f(\mu_1 x^0)$.

According to Figure 2.4

$$E_1 = \frac{OB}{OD}$$

The output increasing efficiency measure shows the ratio between the observed output and the potential output obtained by employing the observed amount of inputs in the frontier function. In Figure 2.3 $E_2 = y^0/y^* = f(\mu x^0)/f(x^0)$. In Figure 2.4,

$$E_2 = \frac{OC}{OD} .$$

The two measures of technical efficiency E_1 and E_2 will generally not coincide, except in the case of linear homogeneity. If $\bar{\varepsilon}$ is an average of elasticity of scale between B' and C' in Figure 2.3, Førsund and Hjalmarsson show that $E_2 = E_1^{\bar{\varepsilon}}$. The ranking of units according to the two measures of technical efficiency coincides if the elasticity of scale is constant or does not pass through the value of 1 in the sample. Thus $E_1 \gtreqless E_2$ when $\bar{\varepsilon} \gtreqless 1$. The choice between the two measures should be determined by the objective. They correspond to the two different approaches in micro economic textbooks: (1) minimising costs for a given output level and (2) maximising output for a given cost level.

Scale efficiency

In studies of industrial structure, the long term potential possibilities of increased efficiency often are the main objective. In many industries, the development of the scale efficiency is of major concern.[1)] A measure of scale efficiency shows how close an observed plant is to the optimal scale. Førsund and Hjalmarsson (1979) derive three different measures of scale efficiency:

The first measure is defined as the relative reduction in input coefficients made possible by producing at optimal scale on the frontier production function with the observed factor proportions. According to Figure 2.4.

$$E_3 = \frac{OA}{OD}$$

This measure is the ratio of a potential input coefficient evaluated at technically optimal scale for the observed input ratios at A' and the corresponding observed input coefficient, at D'. Referring to Figure 2.3, along the rays OD' and OA', the input coefficients, ξ_i $(i = 1,\ldots,n)$

1) See Pratten (1971) and Hjalmarsson (1974).

are constant and equal to the observed, ξ_i^0, and those obtained at optimal scale, ξ_i, respectively. Let a be the slope of OD' and b the slope of OA'. These slopes, being equal to average productivities, may then be used to give the following expressions for E_3

$$E_3 = \frac{\xi_i}{\xi_i^0} = \frac{a}{b} = \frac{y^0}{\hat{y}/\hat{\mu}} = \mu_3$$

where the last expression is arrived at by the simple geometrical relationship

$$\frac{\hat{y}}{\hat{\mu}} = \frac{y^0}{\hat{\mu}3}$$

However, E_3 is not a measure of pure scale efficiency. To obtain such a measure, one has to eliminate the technical inefficiency of the observations by moving each observed unit to the surface of the frontier function. Førsund and Hjalmarsson (1979) define two different measures, E_4 and E_5:

When moving a unit in the horizontal direction, the second measure of scale efficiency, E_4, shows the distance from the transformed isoquant corresponding to y^0 to the optimal scale. In Figure 2.4

$$E_4 = \frac{OA}{OB}$$

When moving a unit in the vertical direction, the third measure of scale efficiency, E_5, shows the distance from the optimal scale to the transformed isoquant corresponding to y*. In Figure 2.4

$$E_5 = OA/OC.$$

E_4 and E_5 give the relative reduction in input coefficients by producing at optimal scale on the frontier function with the observed factor proportions of a plant whose technical inefficiency has been eliminated in two different ways corresponding to the definition of E_1 and E_2, respectively.

According to Figure 2.4

$E_4 = E_3/E_1$, and

$E_5 = E_3/E_2$.

By applying the relationship between E_1 and E_2 discussed above we obtain[1] the following relationship between the scale elasticity and the three different measures of scale efficiency

$$\bar{\varepsilon} = \frac{\ln E_3 - \ln E_5}{\ln E_3 - \ln E_4}$$

where $\bar{\varepsilon}$ is an average scale elasticity between B' and C' in Figure 2.3.

However, there are other approaches in the literature. One approach is suggested by Färe and Lovell (1978). They minimise the value of various distance measures from an observed point to the efficiency frontier. Färe et al (1983) extend this approach to the case of multiple output. However, Russell (1985) has pointed out that it is only the radial measures like Farrell's which can be interpreted within the framework of duality between the production and cost functions.

---

1) See Førsund and Hjalmarsson (1979), pp. 298-299.

2.3.2. Establishing frontier functions

The production theory and efficiency concepts described above serve as a basis for econometric models which can be utilised in the estimation of an intra-industry efficiency distribution. Calculations on the magnitudes and costs of the various types of inefficiency are usually made relative to the efficiency frontiers. In this section we present potential approaches to the building of mathematical and econometric frontier models. The non-frontier approach is initially treated very briefly and so is the estimation of frontier functions via cost functions, because we, in the empirical part of this study, are reporting on the efficiency measurement based on production functions. At the end of the section we briefly discuss empirical applications of the various models.

The frontier functions, relevant in our context, are based on observed performances. In analogy to Salter's concepts, we could call the frontier the best-practice function. We concentrate on the measurement based on statistical information, but note that the engineering approaches to frontier functions are highly relevant, especially for production units on a disaggregated level.[1)]

There are several ways of classifying efficiency measurements with respect to the specifications and estimations used. The frontier may be specified as a

- parametric or
- non-parametric

---

1) See e.g. Eide (1978) for the derivation of an engineering ex ante production function for oil tankers.

function of inputs. In the former case a particular functional form (e.g. Cobb-Douglas or translog) is assumed for the production (or cost) function, while in the non-parametric approaches no functional form is specified. The relationship between observed output and the frontier may be specified as an

- explicit statistical model or
- non-statistical model.

The frontier itself may be specified as either deterministic or stochastic. When we add to these eight permutations the different estimation[1] possibilities, e.g. mathematical programming and regression models, and functional forms from Cobb-Douglas to more flexible functions, we get a whole range of different approaches. Here we concentrate on some of the most relevant of these.

There are, however, empirical studies on efficiency, which are not based on the explicit use of a frontier. In the non-frontier models inefficiency is introduced via varying coefficients[2] or via asymmetry[3], making the use of sophisticated econometric techniques unnecessary, but reducing the information which can be obtained.

One more observation before we come to the main theme, econometric frontier models based on production functions: When using production frontiers, we get information on technical inefficiency but not on

---

1) Management scientists speak of measuring and econometricians of estimating efficiency. Estimating is usually used when we are speaking of a statistical infrence model; see Schmidt (1985).

2) See Lau and Yotopoulos (1971).

3) See Toda (1976) and (1977).

allocative inefficiency, because we only use data on input quantities but not on input prices. The behavioural assumption underlying direct estimation of the cost function is generally cost minimisation, with exogenous output. This requires data on input prices but not on input quantities. The cost frontier yields information on the extra cost of technical and allocative inefficiency but not the separate cost of each, unless further assumptions are made. Førsund and Jansen (1977) have used a deterministic homothetic, and Schmidt and Lovell (1979) a stochastic composed error, Cobb-Douglas cost frontier.

The stochastic frontier model can be extended to allow for separate estimates of technical and allocative inefficiency. The problem is that the allocative inefficiency cannot be calculated except for some special functions e.g. Cobb-Douglas.[1)] Greene (1980) used a system consisting of a deterministic translog cost frontier and associated share equations to gain more flexibility. It is, however, impossible to provide an explicit solution in terms of functional form parameters for the production function corresponding to the translog cost function or vice versa. In this case we cannot exactly distinguish how an inefficiency in one function relates to the corresponding quantity in the other function.

Next we discuss four main approaches to building econometric frontier models based on production functions: deterministic non-parametric, deterministic parametric, deterministic statistical and stochastic frontiers.

---

1) See Schmidt and Lovell (1979).

Deterministic non-parametric frontiers. In a deterministic production frontier model output is assumed to be bounded from above by a deterministic (non-stochastic) production function. Farrell's (1957) derivation of measures for technical, price and overall efficiency considered above in Section 2.3.1. is based on a deterministic non-parametric approach. Farrell formed the free disposal convex hull of the observed input-output ratios by linear programming technique. In Figure 2.2 UU' is the envelope of the observations for all firms. No firm is able to produce a unit of output with an input combination to the southwest of UU', which is thus supported by a subset of the sample with the rest of the sample points above it. The feasibility of the method depends largely on the character and quality of the data available: the frontier is calculated from a supporting subset of observations from the sample and is, therefore, particularly susceptible to extreme observations and measurement errors. The other factor limiting the use of this method is the assumption of linear homogeneity. Farrell and Fieldhouse (1962) generalised this approach to increasing returns to scale by defining a frontier for each chosen level of output; see also Seitz (1970) and (1971). On the other hand, no functional form is imposed on the data.

Deterministic parametric frontiers. Aigner and Chu (1968) used a homogeneous Cobb-Douglas production frontier with all observations on or below the frontier

$$\begin{aligned} \ln y &= \ln f(x) - u \\ &= \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln x_i - u, \qquad u \geq 0 \end{aligned}$$

The error term u is one-sided and forces $y \leq f(x)$. The elements of the parameter vector $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_n)'$ may be estimated either by linear or quadratic programming. The "estimates" obtained through mathematical programming have no statistical properties, such as t-ratios and standard errors. This fact, which makes the appraisal of the results obtained more difficult, is due to the absence of stochastic assumptions made about the regressors or the disturbance; without some statistical assumptions of the distribution of the error terms no inferential results can be obtained. The method is sensitive to extreme observations, the same disadvantage found in the case of the non-parametric frontier.

The advantage of the parametric approach over the non-parametric one is that it characterises the frontier technology in a simple mathematical form. A further advantage is also the possibility to handle non-constant returns to scale.

Deterministic statistical frontiers with an explicit efficiency distribution. A step further in comparison with the mathematical programming approach is to introduce an explicit efficiency distribution for u and derive maximum likelihood estimates of the parameters inside the deterministic framework. This was first suggested by Afriat (1972) who pointed to the Beta distribution as the most general distribution satisfying the natural requirements of such an efficiency distribution[1].

The problem with maximum likelihood estimation is that the underlying conditions for the application of maximum likelihood are not met. Due to the on-or-below frontier constraints, the range of the stochastic

---

1) For estimation of the parameters of the beta distribution by maximum likelihood method see Vartia (1973).

output variable depends on the parameters to be estimated, which makes the properties of ML estimators uncertain.[1)]

The basic model may be written as follows:

$$y = f(x)\, e^{-u}$$

$$\ln y = \ln\,[f(x)] - u, \qquad u \geq 0.$$

In a Cobb-Douglas case the ln [f(x)] is linear with respect to parameters and inputs. We may assume that the observations on u are independently and identically distributed and that x is exogenous or independent of u. Based on these assumptions, different distributions for u could be specified. The choice of a distribution for u (or equivalently for e^{-u}) is crucial. Afriat (1972) used a two-parameter Beta distribution for e^{-u} but there is no good a priori argument for any particular distribution, which is a source of problems in empirical studies.[2)]

This approach is followed up in an empirical application by Broeck et al (1980) and Førsund and Hjalmarsson (1984). They specified a one parameter efficiency distribution of the following type:

$$h(u) = (1 + a)\, e^{-(1+a)\, u}$$

1) See Schmidt (1976).

2) A promising possibility might be to consider an adaptive estimator proposed by Manski (1984). This would estimate both the parameters and distribution of u. See also Schmidt (1985).

Another type of a 'probabilistic' frontier is just a deterministic frontier computed from a subset of the original sample. Hypothesis testing is impossible because it is computed rather than estimated. So the problem remains the same as when using the deterministic frontiers; see Timmer (1971).

Stochastic frontiers. In the three approaches presented briefly above, all firms share a common family of production, cost and profit frontiers and all variation in firm performance is attributed to variation in firm efficiencies relative to the common family of frontiers. But we can specify a stochastic production frontier, where the output of each firm is bounded from above by a frontier that is stochastic in the sense that its placement is allowed to vary randomly across firms. From an economic standpoint this method permits firms to be technically inefficient relative to their own frontier rather than to some sample norm. The advantage is that in the stochastic frontier model the error term is composed of two parts: a symmetric or normal component capturing randomness outside the control of the firm and a one-sided (non-positive) component capturing efficiency under the control of the firm.

We can write the model as

$$y = f(x) \exp(v - u).$$

$f(x) \exp(v)$ is the stochastic production frontier. v has some symmetric distribution to capture the random effects of measurement error and exogenous shocks not in the control of the firm (e.g. weather, strikes etc.) which cause the placement of the determintistic kernel $f(x)$ to vary across firms. The other component u is one-sided and it captures the technical inefficiency, resulting from failure to produce the maximum possible output from the use of a given set of inputs, relative to the stochastic production frontier. The estimation of the frontier is a statistical problem and the normal statistical inferences are possible in principle. This model was introduced by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). It is a natural extension of earlier work on stochastic frontiers emphasising skewness of the error distribution.[1)]

To introduce specific probability distributions for the disturbance terms[2)], the maximum likelihood method can be applied. Since the constraint set no longer applies, the ML estimators have their standard properties. However, the general stochastic, composed error model has a serious disadvantage: it is difficult to decompose individual residuals into their two components. It is possible, however, to obtain conditional estimates of technical inefficiency for each firm from the estimated residuals. In 1982 Jondrow et al. developed a method to estimate technical inefficiency from the residuals which include both technical inefficiency (u) and general statistical noise (v) by considering the expected value of u, conditional on (v-u).

Otherwise we cannot estimate technical inefficiency by observation, only the average level of efficiency for the industry can be obtained. This implies that intra-industry analysis of the firms is precluded, but inter-industry analysis can be performed with respect to average efficiency levels.

---

1) See e.g. Aigner, Amemiya and Poirier (1976).

2) E.g. Aigner et al. (1977) and Meeusen and van den Broeck (1977a) used the lognormal distribution for the pure random term and truncated normal (Aigner et al. (1977)) and exponential (Meeusen and van den Broeck (1972b)) for the inefficiency term.

Panel data. Some of the disadvantages connected to the usage of the stochastic, composed error models can be avoided by using panel data. With panel data it is possible to introduce further assumptions on the development of firm specific efficiency over time. Having data on a cross-section of firms, each observed for a number of time periods, it is possible to test whether the efficiency variable is constant or not over time for each firm.[1)] At the one extreme, the error terms reflecting inefficiency can be taken to be independent over time, in which case the panel nature of the data is irrelevant. At the other extreme, we can assume that firm inefficiency is constant over time, in which case the panel data approach is relevant. Unchanging inefficiency over time is a powerful assumption, since it allows one to remedy certain serious problems of stochastic frontier models, but at the cost of loss of realism.

Schmidt (1985) points at three problems which can be avoided by using panel data. Firstly, one problem with stochastic frontier models is the strength of the distributional assumptions on which they rely. Using panel data such strong assumptions are unnecessary.[2)] Secondly, inefficiency and input levels are assumed to be independent in the stochastic frontier models, especially in the single equation model. With panel data no such assumptions is necessary. The fixed-effects estimator does not make or require any assumption of independence between the effects and the explanatory variables.[3)] The third problem with stochastic frontiers models is the question of separation of

---

1) See e.g. Pitt and Lee (1981).

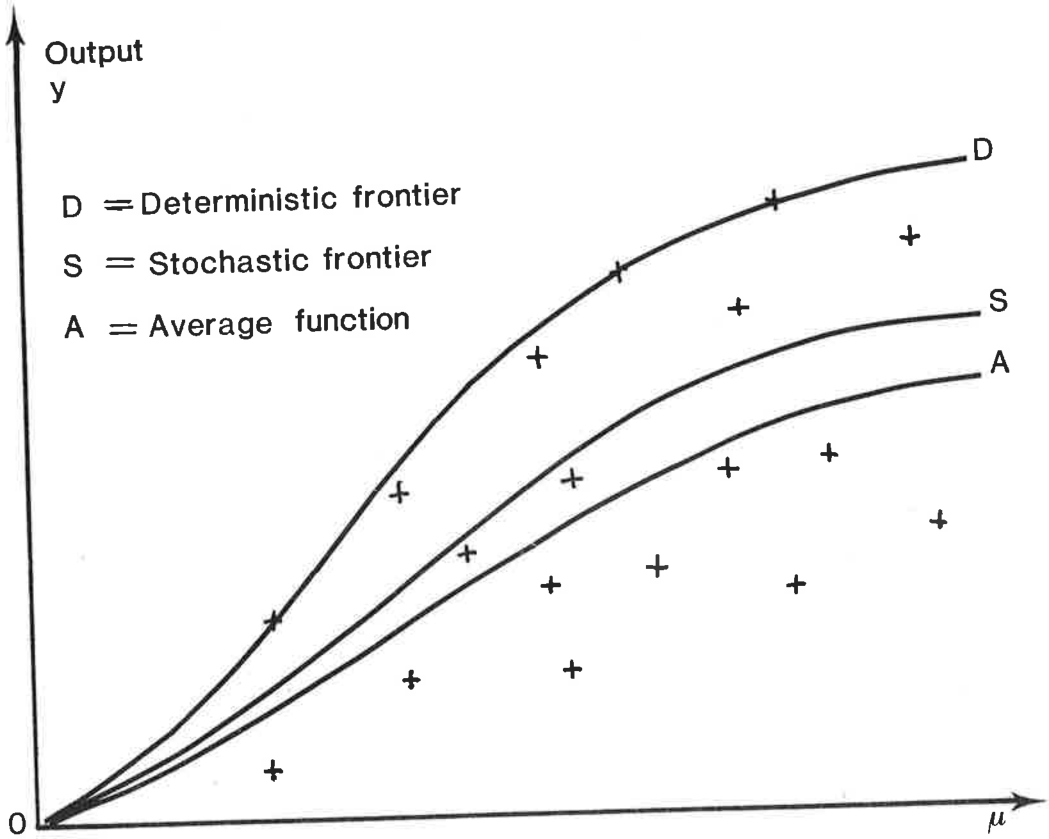
2) See Schmidt and Sickles (1984).

3) However, if we make such an assumption, more efficient estimation is possible and the assumption is testable.

technical inefficiency from statistical noise. With panel data it is possible to get better estimates of technical inefficiency when we have a number of time series observations instead of one cross section.

In Figure 2.5 we compare the deterministic and stochastic frontiers and the "average" function estimated from the observed data. In a deterministic frontier model all unexplained randomness is taken to be inefficiency and so it shows higher inefficiency than stochastic frontier models.[1] The problem of outliers is obvious and has been tackled by

Figure 2.5. The position of deterministic and stochastic frontier production functions and the average function.



1) Cowing, Reifschneider and Stevenson (1983) compared the results of applying a non-frontier (normal distribution) model, a deterministic (gamma) frontier model and two different stochastic frontier (normal-exponential and normal-half normal) models to the same data. The average inefficiency is much higher for the deterministic frontier model than for either stochastic frontier model.

Timmer (1971) and many others. Timmer (1971) estimated a so-called probabilistic frontier by removing all the efficient observations on the frontier and then recomputing a deterministic frontier without these observations. The solution seems to be too arbitrary. Broeck, Førsund, Hjalmarsson and Meeusen (1980) carried out a sensitivity test by removing the unit with the highest shadow price on or below the frontier constraint for each cross-section sample. We will test the sensitivity of our main results in a similar way in Section 3.3.1.3.

The main approaches for establishing frontier production models have been discussed above. Recent theoretical and empirical emphasis lies mainly on more sophisticated stochastic models and the models using panel data[1)], their statistical properties and on the estimation of both technical and allocative inefficiency relative to stochastic production and cost frontiers.[2)] Simultaneously, more flexible functional forms are introduced.[3)] Broeck, Førsund, Hjalmarsson and Meeusen (1980) have compared the various frontier approaches using data on Swedish milk processing plants. The results differ, depending on the choice of model. In Table 2.1 some of the most typical approaches and empirical applications are summarised.[4)]

---

1) E.g. Pitt and Lee (1981), Schmidt and Sickles (1984), Bauer (1984), Melfi (1984) and Kumbhakar (1985).

2) E.g. Schmidt and Lovell (1979) provide a generalisation of a stochastic model to allow for a correlation between technical and allocative efficiencies within a firm.

3) E.g. Greene (1980) used the translog frontier specification to characterise the firm's production technology. See also Kopp (1981), and von Maltzan (1978).

4) Schmidt (1985) gives an overview of the empirical experience with frontier techniques.

Table 2.1: A survey of basic approaches for the estimation of explicit frontier production functions.

TYPES OF FRONTIERS / STO-CHASTIC SPECIFICATIONS	DETERMINISTIC FRONTIER: WHOLE SAMPLE ON OR BELOW THE FRONTIER	STOCHASTIC FRONTIER: NO ON OR BELOW THE FRONTIER RESTRICTIONS ON OBSERVATIONS
EFFICIENCY FIXED	PROGRAMMING METHODS - Aigner & Chu 1968 - Timmer 1971 - Førsund & Hjalmarsson 1979	PANEL DATA - Schmidt & Sickles 1984 - Kumbhakar 1985(a)
EFFICIENCY RANDOM BUT NO SPECIFIC DISTRIBUTIONAL ASSUMPTION	CORRECTED OLS - Richmond 1974	(Not possible)
EFFICIENCY RANDOM WITH EXPLICIT DISTRIBUTIONAL ASSUMPTION	MAXIMUM LIKELIHOOD ESTIMATES - Afriat 1972 - Schmidt 1976 - Greene 1980 - Broek et al 1980	DIFFERENT WEIGHTS ON POSITIVE AND NEGATIVE RESIDUALS - Aigner et al 1976 COMPOSED ERROR - Aigner et al 1977 - Meeusen & Broeck 1977 - Førsund et al 1980
EFFICIENCY RANDOM: PANEL DATA ANALYSIS		SINGLE EQUATION - Pitt & Lee 1981 - Schmidt & Sickles 1984 MULTIPLE SIMULTANEOUS EQUATIONS (OPTIMIZING FRAMEWORK) - Bauer 1984 - Melfi 1984 - Kumbhakar 1985(b)

2.3.3. Further characterisation of technical progress

In this section we expand the analysis to the characteristics of technical change. There are several possible ways of measuring the impact of technical change in our framework. For our purposes the approach outlined by Førsund and Hjalmarsson (1979) is the most suitable one because it derives directly from Farrell's (1957) and Salter's (1960) measures.

Salter (1960) introduced, limiting his study to two factors and assuming constant returns to scale, three measures of technical advance[1]:

1 The rate of technical advance measured by the relative change in total unit cost for constant input prices and output level; This measure is designated by T.

2 Labour, or capital saving bias measured by the relative change in the optimal (cost minimising) factor proportion for constant input prices; This measure is designated by D_{ij}.

3 The relative change in the elasticity of substitution.

Førsund and Hjalmarsson (1979) generalised the first two measures to n factors in the case of non-homogeneous production functions in the following way:

The rate of technical advance measure can be split up into two main components:

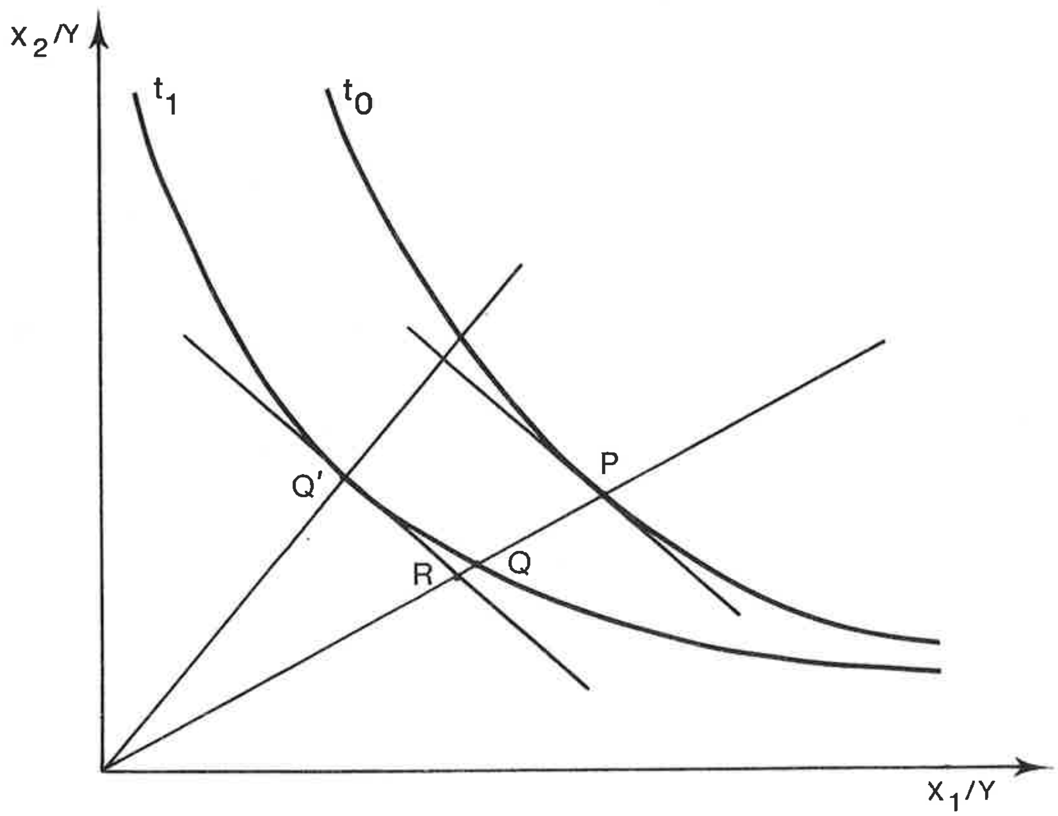
T_1 = the reduction in unit costs due to the movement along a factor ray

T_2 = the reduction in unit costs due to the movement along the next period's efficiency frontier.

1) See Summa (1979), pp. 31-36.

T_1 is called the <u>proportional technical advance</u> and T_2 the <u>factor bias advance</u>. Both have similarities with Farrell's efficiency concepts, which can be illustrated by the two-factor case using Figure 2.6. P is the point of reference on the efficiency frontier for the starting period t_0. Q' is the point on the efficiency frontier for a later period t_1 with the same factor prices. A measure of technical advance analogous to Farrell's, assuming cost minimisation, is the relative change in unit cost from P to Q', i.e. the unit cost reduction possible when choosing techniques from two different ex ante functions for constant factor prices and realising the optimal scale. This change OR/OP is analogous to Farrell's overall efficiency measure when P is an observed unit and t_1 the efficiency frontier. Salter's technical advance measure can be decomposed multiplicatively into proportional advance, T_1, (OQ/OP) and bias advance, T_2, (OR/OQ) analogous to Farrell's technical efficiency and price efficiency, respectively.

Figure 2.6. The generalised Salter measure of technical advance and its components.



T_1 shows the relative reduction in unit cost due to a movement along a factor ray, and correspondingly T_2, shows the relative reduction in unit cost due to the movement along the period t_1 efficiency frontier generated by biased technical change. Thus $T = T_1 \cdot T_2$. In the case of a homothetic production function, the unit cost reduction due to the movement along a factor ray T_1 can be further decomposed into a reduction in unit cost due to a change in optimal scale (OS), a cost reduction due to Hicks-neutral technical progress (H), and a cost reduction due to factor bias technical change for constant factor ratio (B) and T_1 = OSxHxB.

Let us now look at the bias of technical advance. Salter's measure of factor bias shows the change in the optimal factor ratio for a pair of inputs given constant factor prices between time points t_0 and t_1. The general version of the Salter bias measure is

$$D_{ik} = \frac{X_{i,\ t_1}}{X_{k,\ t_1}} \Big/ \frac{X_{i,\ t_0}}{X_{k,t_0}} = \frac{X_{i,\ t_1}}{X_{i,\ t_0}} \cdot \frac{X_{k,\ t_0}}{X_{k,\ t_1}}$$

for constant input prices and output level.

In a case with more than two inputs, this measure becomes a relative concept, dependent on the factor pair under consideration. If we want a common basis for classifying the nature of the bias, one possibility is to look at changes in the cost shares, S_i, for constant input prices and output level. This measure was proposed by Binswanger (1974).

2.4. The short-run industry production function (SRIPF)

2.4.1. Introduction

In order to develop a comprehensive analysis of intra-industrial efficiency, technical progress and structural change, information about both the ex ante micro function and the short-run industry production function (SRIPF) is required in the putty-clay model introduced in Section 2.1. However, performing such a complete analysis requires very detailed information about technical relationships. In this study we emphasise the links between these two production functions and, as far as we know, are the first to estimate both the frontier and the SRIPF in the same study. We first describe the properties and the uses of the SRIPF in Section 2.4.1. and then define the concept more formally in Section 2.4.2. Finally, we show the SRIPF's potential in deriving several concepts of technical and structural change.

The ex ante function may be derived from engineering knowledge[1)] or estimated as a frontier production function, as we have discussed in Section 2.3.2. The concepts of the ex ante and the SRIPF are complementary. Analytically the connection between a series of the SRIPFs over time goes through the ex ante production functions of the micro units with the fixed factors as variables. The SRIPF reflects both the history of ex ante functions over time and the actual choices made from these ex ante functions. Production at any point of time must be compatible with the SRIPF. It captures well the dynamics of the industry since both the properties of new investments as well as the scrapping of old capacity form an essential part of the development of the SRIPF.

---

1) See e.g. Eide (1979).

The basic idea of the SRIPF can be summarised as follows. We consider an industry producing a homogeneous output and comprising a certain number of firms. When investing in new equipment a firm chooses technology according to the ex ante production function which may be assumed to exhibit all traditional neoclassical properties. After the investment has been carried out a new production unit appears. The production possibilities of this unit are described by the ex post production function at the micro level. This is assumed to be a limitational law (fixed proportions production function) and in addition there exists a maximum production capacity for this unit. Each plant is thus characterised by fixed production coefficients with regard to current inputs and the presence of fixed factors in the form of existing capital. Fixed capital only determines the capacity of the individual micro-units and does not appear as an explicit variable in the ex post function. Furthermore, it is assumed that there are no costs associated with the utilisation of the fixed factors in the short run. Aggregating in an efficient way (as described in more detail in Section 2.4.2) all existing production units, characterised by their ex post production functions, yields the SRIPF.

The putty-clay structure introduced by Johansen and Salter is the key assumption in establishing the SRIPF. The concept of the short-run production function is based on an aggregation of micro production functions. This aggregation process is based on maximising industry output for a given level of inputs. The production function obtained is non-parametric in the case of a discrete distribution of micro units. It allows a richer numerical description of the development of the industry's characteristics than conventional approaches based on the estimation of an "average" production function for the industry on the basis of the notion of a representative firm.

The maximising approach applied here when constructing the short-run function corresponds to the basic definition of a production function, in which an industry is regarded as one production unit, as opposed to the traditionally estimated "average function" for an industry, in which it is assumed that all micro units have the same underlying production technology, except for a random error term. According to the basic definition of the production function in the pure theory of production[1)], the production function in the technical sense provides the maximum amount of output attainable with given amounts of inputs. The SRIPF fulfills this basic definition.

The SRIPF explicitly recognises that the technologies and capacities of individual micro units do differ, and utilises all these individual micro-technologies when establishing the relationship between the aggregate industry output and current inputs by explicit optimisation.

The aggregation is performed, as noted earlier, by maximising the industry's total production for given amounts of total inputs. This also means that the industry's total production costs are minimised for any factor price ratio and any level of production, assuming that all units of production face the same prices. Thus the approach also implies a non-parametric minimum cost function from which average costs and marginal costs may be calculated at different levels of output.

From the SRIPF several concepts may be derived which reveal different pro-

perties of the function and provide valuable insights into the structure

---

1) See e.g. Frisch (1965).

of an industry. The structure of an industry is here characterised by the shape and location of the substitution region (economically relevant region) and the shape and spacing of the isoquants. The latter depends on the distribution of the technical properties of the micro units. The structure of an industry may be summarised in the SRIPF by various concepts of elasticities, such as the elasticities of scale, cost and substitution. In the SRIPF these parameters are variables and depend in general on the point of calculation in the substitution region.

The SRIPF shows how to organise the industry in the most efficient way when the degree of capacity utilisation and current factor prices vary and given that all units face the same factor and output prices. Besides this normative approach, often relevant to industrial policy and managerial decision making at the industry level, there is a positive interpretation too. The SRIPF may also describe industry behaviour under decentralised decision making in perfect competition.[1)]

2.4.2. Definition of the short-run industry production function

We now consider an industry consisting of a certain number of micro-units and define the short-run industry production function following Johansen (1972) and Førsund and Hjalmarsson (1984). The short-run production function of the industry as a whole is established on the basis of an ex ante production function, determining the full capacity values y, $\bar{x}_i$ $(i = 1,\ldots,n)$ of homogeneous output y and the current

---

1) See Førsund-Gaunitz-Hjalmarsson-Wibe (1980), pp. 118-119.

inputs $\bar{x}_i$. For the ex post function at the micro level, according to Johansen (1972), a limitational law (Leontief technology) is assumed to hold:

$$y = \text{Min}\left[\frac{x_1}{\xi_1},\dots,\frac{x_n}{\xi_n},\bar{y}\right] \qquad (1)$$

where the input coefficients $\xi_i = \frac{\bar{x}_i}{\bar{y}}$ $(i = 1,\dots,n)$ are constant, i.e. independent of the rate of capacity utilisation.

In deriving the SRIPF we allow all the plants to have different input coefficients and also different production capacities. All units have the simple structure of (1) and the input coefficients are estimated by the observed coefficients using engineering data on plant capacities.

The short-run industry production function $Y = F(X_1,\dots,X_n)$ is obtained by solving the following problem:

$$\text{Max } Y = \sum_{j=1}^{m} y^j \qquad \text{subject to} \qquad (2a)$$

$$\sum_{j=1}^{M} \xi_i^j\, y^j \leq X_i, \qquad i = 1,\dots,n \qquad (2b)$$

$$y^j \in [0, y^j], \qquad j = 1,\dots,m \qquad (2c)$$

where Y denotes output and $X_1,\dots,X_n$ current inputs for the industry as a whole, and where $j = 1,\dots,m$ refers to plants with a capacity of y^j. Since, for our purposes, we are only interested in the economic region, it is natural to assume free disposal of inputs as expressed by equation (2b).

In order to get a better interpretation of the condition for utilising units we study the dual problem of the problem (2) following Førsund and Hjalmarsson (1984). Let the dual variables $q_1, q_2, \ldots, q_n$ correspond to the restrictions on current inputs in equations (2 b) and dual variables $r^1, r^2, \ldots, r^m$ correspond to capacity limitations in equations (2 c) respectively. The dual problem is to minimise

$$\sum_i q_i X_i + \sum_j r^j \bar{y}^j \quad \text{subject to} \tag{3 a}$$

$$\sum_i q_i \xi_i^j + r^j \geq 1 \quad j = 1, \ldots, m . \tag{3 b}$$

The necessary first order conditions yield:

$$(4) \quad 1 - \sum_{i=1} q_i \xi_i^j \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} 0 \quad \text{when} \tag{4}$$

$$\begin{cases} y^j = \bar{y}^j \\ y^j \in [0, \bar{y}^j] & j = 1, \ldots, m \\ y^j = 0 \end{cases}$$

The variables, $q_1, \ldots, q_n$, are shadow prices of the current inputs with dimension output per unit of input. It follows directly then, that $q_1, \ldots, q_n$ represent the marginal productivities of the inputs of the industry function. Whether a production unit is to be operated or not is decided by whether current operation "costs" (dimensionless), calculated at these shadow prices, would be lower than or exceed unity. This corresponds to utilising units with non-negative quasi-rents. An equality sign in (4) defines the zero quasi-rent line in the input coefficient space, thus giving the boundary of utilisation of the set of production units.

When operating costs equal unity we have a marginal production unit in the sense that it may or may not be operated in the optimal solution.[1)]

Johansen's illustration of the SRIPF for three production units is shown in Figure 2.7, where the capacities are indicated in the following way: The coordinates of E ($\xi_1^1 y^1$, $\xi_2^1 y^1$) indicate the inputs of factors 1 and 2 when production unit 1 is used to full capacity. The points F and A indicate full capacity utilisation of production units 2 and 3, respectively. For total factor amounts corresponding to X_1, X_2-points within the parallelogram[2)] OFDEO production is possible and efficient with utilisation of production units 1 and 2.[3)] The isoquants in this region are parallel to the line between (ξ_1^1, ξ_2^1) and (ξ_1^2, ξ_2^2). For factor amounts corresponding to points within the parallelogram OABFO production is possible and efficient with utilisation of production units 2 and 3 and the isoquants are parallel to the line between (ξ_1^2, ξ_2^2) and (ξ_1^3, ξ_2^3).

For total factor combinations corresponding to points in the region FBCDF it is necessary to use all three production units. The unit 2, superior to combinations of the processes of units 1 and 3, will always be used to full capacity in this region, whereas units 1 and 3 are combined to produce output in excess of the full capacity output of unit 2. The isoquant slopes are determined by the input coefficients of units 1 and 3. The point C determines the total capacity of the industry and at this point all units are utilised to full capacity.

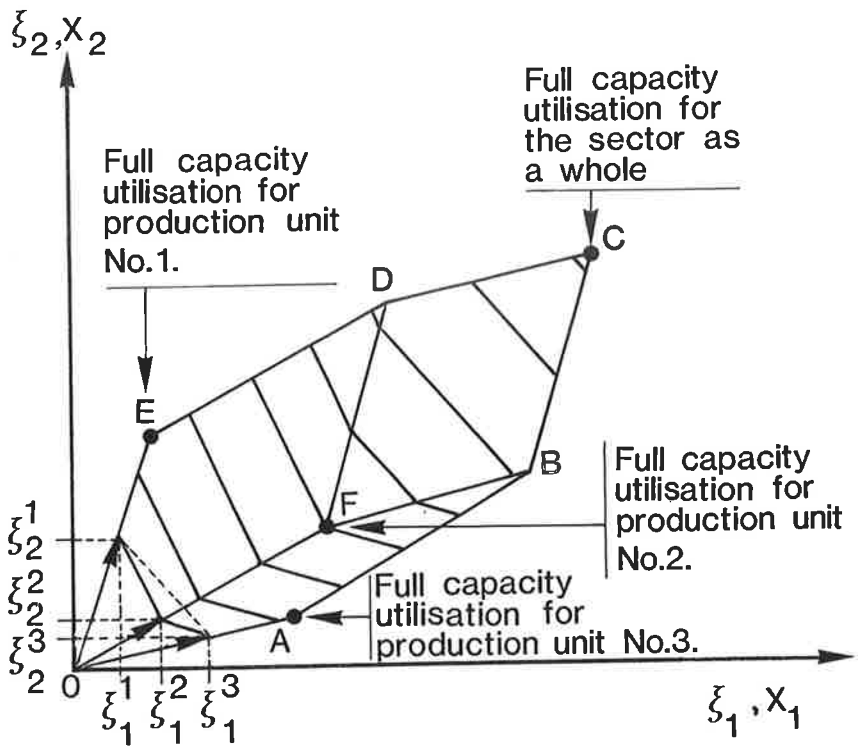
---

1) See Johansen (1972) pp. 13-19 for a detailed exposition. See also Bosworth (1976) pp. 11-13.

2) Such a parallelogram is also called an activity region.

3) For further explanation see Førsund and Hjalmarsson (1984), Chapter 5.

Figure 2.7: Short-run industry production function for three units (Source: Johansen (1972), p. 17)



When the input coefficients are assumed to be constant, the output optimisation problem is a linear programming one. If the coefficients are functions of the capacity utilisation we have a non-linear programming problem.

Due to the assumed linear structure of the problem (2a-c), the isoquants will be piece-wise linear in the two-factor case considered here. In principle, the short-run function (2) can be derived by solving a prolific number of LP-problems, but in practice this is not a feasible procedure if one's interest is in establishing a reasonable number of isoquants.[1)] Instead we use an algorithm, developed by Førsund and

---

1) A limited number of isoclines is readily obtained by a simple ranking of the micro units according to unit production costs for given input prices. Johansen (1972), W. Hildenbrand (1981) and K. Hildenbrand (1982) have used such a cost minimisation procedure.

Hjalmarsson[1], which yields, for the two-factor case, a complete description of the isoquants by locating all the corner points geometrically, providing the entire set of isoclines.

The short-run industry production function, defined above, is non-parametric. The question of how the function should be represented now arises. This must depend on the use to which the function is to be put. In order to analyse long-run technical progress, we need the complete representation of each isoquant of the set found suitable for analysing the three aspects:

1 factor bias
2 change in productivity and
3 change in substitution properties

Changes in the capacity distribution are generated by several factors besides technical progress as expressed by changes in the ex ante functions.[2] One might, therefore, expect that changes in the short-run function would be more complicated and less likely to be captured by the limited number of parameters of an analytical production function. Moreover, it has been shown (see examples e.g. in Sato (1975) and Hildenbrand (1981)) that, given the basic hypothesis of the specification, the parametric functional form implies strong assumptions on the distribution of technical information over micro-production units, i.e. there is a relationship between families of parametric production functions and distributional forms of micro-units (see Seierstad (1985)).

---

1) For a description of this algorithm, see Førsund and Hjalmarsson (1984).

2) See Johansen (1972), p. 29-33 for a discussion of such factors.

In the empirical analysis of this study we do not, either explicitly or implicitly, make any assumptions regarding parametric functional forms of industry production functions, or equivalently, distribution functions of micro-units. Such an assumption is implicitly made in the conventional econometric analysis of production functions estimation. That is why we have chosen the non-parametric approach. The non-parametric SRIPF and derived concepts may be illustrated numerically. A numerical presentation of various complicated functions is, however, very impractical. The results are, therefore, usually visualised by means of plottings of the substitution regions and isoquants for the different output levels so that the changes are directly observable (see e.g. Figures 3.4 and 3.5 in Section 3.4.2.).

2.4.3. Further characteristics of the SRIPF

The SRIPF shows the actual, chosen production possibilities of the industry and it is changed by the investment in new technologies and the scrapping of old capacity. From the SRIPF several concepts may be derived which reveal different properties of the function and provide valuable insights into the structure of an industry. First we present supplementary ways of illuminating the industry structure in the SRIPF-framework. Then we apply this production function approach more specifically in the analysis of scale and related cost properties and in the alternative ways of measuring technical change.

In addition to isoquant maps there are other ways of presenting the SRIPF and illuminating the nature of intra-industrial technical change. The set of ex post production functions in the input coefficient space is called the capacity distribution. This can be represented by a diagram in which each production unit is characterised by its input coefficients and capacity (for an example see Figure 3.4). Actually the SRIPF is constructed on the basis of such a capacity distribution. Transforming the SRIPF to the input coefficient space yields the capacity region (see Figure 3.16), which represents the feasible input coefficients of the industry production function as a whole while the capacity distribution shows the dispersion of individual units.

In an analogous way Førsund and Hjalmarsson have portrayed the complete efficient combinations of the micro units[1]. Starting at zero industry production and expanding this to full capacity utilisation the activity regions, like the parallelograms OFDE, OABF and FBCD in Figure 2.7, are formed by adding micro units in accordance with the requirement that maximum industry output is obtained at each point in the substitution region. Each parallelogram is formed by combining two units. Within the parallelogaram the utilisation rate of a marginal unit is between zero and one. The activity regions representation contains the complete set of all possible isoclines, since each line segment of the parallelogram represents the locus of isoquant corners.

Using activity regions we can follow each individual unit's utilisation as a function of the industry's capacity utilisation. Each unit is moved in parallel shifts in a strip-like fashion from one boundary of the substitution region to the other. Førsund and Hjalmarsson call the graph of this kind of movement of units partial or marginal utilisation strips. In a similar way the various technologies employed in an industry can be analysed. In the latter case we call them technology strips.

---

1) Førsund and Hjalmarsson (1984), Chapters 5 and 8.

Substitution properties. There is, of course, no direct substitution between inputs of various micro-units due to the assumption of fixed coefficients. Firms with several micro units can, however, choose between different types of capital equipment and in this way obtain different factor proportions (or input coefficients) in the firm as a whole. This dispersion of technology between micro units makes substitution at the industry level possible, since a given amount of output can be produced with different combinations of micro units. Therefore, the substitution properties at the different points of the isoquant map are variable.

The elasticity of scale function in the SRIPF context is in principle defined in the same way as in the traditional theory of production; see e.g. Frisch (1965), Krelle (1969). However, due to the piece-wise linear nature of the SRIPF concepts based on derivatives are not defined uniquely at corner points. Along an expansion path there are two values for the elasticity of scale: a left hand one and another to the right. From the construction of the short-run function it follows that the maximum value of the scale elasticity is one. When we move outwards along the expansion paths, the elasticity in general decreases but not monotonically.

The elasticity of costs with respect to output is, as usual, defined as the ratio between marginal costs and average costs. This concept is well defined here. In the case of a continuous production function the inverse of this ratio is equal to the elasticity of scale. In general, the cost elasticity will be between the inverse of the value of the left hand and the right hand scale elasticities. Obviously the minimum value of the elasticity of cost is one but has to increase with some variation expanding the industry capacity.

The variation in scale and cost elasticities is due to the fact that changes in elasticities occur when moving from one production unit to another along the expansion path and there are differences with respect to input coefficients between these. The graphs of average and marginal cost curves along an expansion path provide us with a comprehensive picture of the variable cost structures for each output level. If the differences between the units are large, there are jumps in the elasticities. Moreover, the scale and cost elasticities give information about the structure of the industry, particularly the spacing of isoquants. Thus it allows a deeper insight into the heterogeneity of the cost structure of the micro-units.

2.4.4. Measuring technical change

Earlier, in Section 2.3.3, we presented a measure of technical change in the spirit of Salter. In the SRIPF context the measurement of technical change may be performed along the lines of Salter too. The first feature of technical change which is important is the rate of movement of the isoquants of the production function towards the origin. The extent of technical advance from one period to another in the SRIPF is defined and measured by the relative change in total unit costs between two points in time, t_1 and t_0, $t_1>t_0$, at a certain output level y^* and at chosen constant factor prices w^*:

$$(1) \qquad T = (c_{t_1}(y^*,w^*)/y^*)/c_{t_0}(y^*,w^*)/y^*$$

Here y^* is the chosen output level and c_t the minimised costs at time t_k, k=0,1, at production level y^* and prices w^*.

The second important feature of technical advance relates to the bias towards uneven factor saving. Salter's measure of factor bias is defined as the relative change in the values of the ratios of the cost minimising factor demand functions at a given output level and at constant factor prices w*:

(2) $$D_{ij} = [x_{i,t}\,(y^*,w^*)/x_{j,t}\,(y^*,w^*)]/[x_{i,t}\,(y,w^*)/x_{j,t}\,(y^*,w^*)]$$

where $i = j$, $i,j = 1,\ldots,n$.

The bias-measure shows the uneven relative change in the location of the expansion paths corresponding to prices w* due to technical change.

Technical change may be classified by factor saving bias. If in (2) D_{ij} is greater (less) than one this means that technical change is factor-i-using (saving) relative to factor j. Biases and technical change shift the location of the substitution region in an uneven way, changing the spacing and shape of isoquants.

Essentially the procedure of measuring technical change is analogous to the index number problem, for it involves asking what changes in the unit cost function (or production function) would take place if relative prices and output were constant. In this way substitution type changes in technique may be eliminated at constant output levels and the characteristics of technical advance described by reference to techniques which differ only by shifts in unit cost or production function from one period to another. In an industry with rapid technical advance, there would be large differences in the position of corresponding isoquant levels; in a technically stagnant industry, the isoquants would be stationary.

Technical change has an impact on production cost. The cost saving nature of technical change may be decomposed into several parts. Here we restrict ourselves to its effect on saving from the price side of production.

The relative change in the price of production cost between two time points t_1 and t_0, $t_1>t_0$ - the true price index of production at production level y* and prices w_{t_1} (in Karko's (1986) terms) - may be decomposed into two parts:

$$(3) \qquad \frac{c_{t_1}(y^*,w_{t_1})}{c_{t_0}(y^*,w_{t_0})} = \frac{c_{t_1}(y^*,w_{t_1})}{c_{t_0}(y^*,w_{t_1})} \frac{c_{t_0}(y^*,w_{t_1})}{c_{t_0}(y^*,w_{t_0})}$$

The first part of this decomposition is just the above mentioned Salter measure of technical change at production level y* and prices w_{t_1}. The second term is a conventional economic price index of production cost at production level y* and prices w_{t_1} measured within the base techniques with the corresponding cost function c_{t_0}. The true index is measured as the production cost ratio between both techniques at their isoquants level y*. The index shows how much less it would cost to utilise the optimal input bundle at prices w_{t_1} to produce a given amount of product y*, by using the techniques t_1 compared to the input bundle needed in techniques t_0 at prices w_{t_0} to produce the same amount y* of product. The economic index measures the same, but in the base techniques t_0. So it standardises true cost comparisons to yield a measure of technical deviation in economic price terms between the true and hypothetical production situation.

Changing the base technique changes the economic index and thus also the measure of technical change. So price and technical bases also affect the measures of technical change.[1)] On the other hand, Karko (1986) has shown that the true price index also decomposes into relative changes in scale and marginal price. The former term is called the true price index of scale and the latter the true price index of the marginal cost. The true indices may be further decomposed into two components. The first component is in both terms affected by technical change only and measured at production level y* with prices w_{t_1} while the second component in both terms is economic scale and marginal cost indices between the price situations w_{t_1}, and w_t calculated in base techniques along the isoquant level y*.

The true volume index of production costs may also be decomposed in the same spirit as that of the true price index. These decompositions depend also on the basis of the comparisons. The true volume index of production costs at prices w* and production level $y^*_{t_0}$ in base techniques t_0 may be presented as follows[2)]

$$(4) \quad \frac{c_{t_1}(y^*_{t_1}, w^*)}{c_{t_0}(y^*_{t_0}, w^*)} = \frac{c_{t_1}(y^*_{t_1}, w^*)}{c_{t_0}(y^*_{t_1}, w^*)} \, \frac{c_{t_0}(y^*_{t_1}, w^*)}{c_{t_0}(y^*_{t_0}, w^*)}$$

---

1) This price index coincides with the concept of conventional economic (or functional) price-index of Konüs-type; cf. e.g. Diewert (1981), Samuelson and Swamy (1974).

2) In the above decomposition the measure of technical change may be considered as Paasche-type. It is also possible to define Laspeyres type of measures by choosing prices w_{t_0} and changing the base technics in economic index number to be t_1.

By choosing in (3) $y^* = y^*_{t_1}$ and in (4) $w^* = w^*_{t_1}$ we get the same technical change measures for both decompositions (3) and (4). Under these circumstances the optimal value of production cost to produce with techniques t_1 at output level $y^*_{t_1}$ at prices $w^*_{t_1}$ compared to the optimal cost of producing with techniques t_0 amount $y^*_{t_0}$ at prices $w^*_{t_0}$ may be presented e.g. as follows

$$(5)\qquad \frac{c_{t_1}(y^*_{t_1},w^*_{t_1})}{c_{t_0}(y^*_{t_0},w^*_{t_0})} = \frac{c_{t_1}(y^*_{t_1},w^*_{t_1})}{c_{t_0}(y^*_{t_1},w^*_{t_1})}\,\frac{c_{t_0}(y^*_{t_1},w^*_{t_1})}{c_{t_0}(y^*_{t_1},w^*_{t_0})}\,\frac{c_{t_0}(y^*_{t_1},w^*_{t_0})}{c_{t_0}(y^*_{t_0},w^*_{t_0})}$$

This may be considered an example of an extension of Fisher's weak factor reversal test in the conventional theory of economic index numbers. Thus, in general, deflating the cost ratio by the true price index of production cost yields the economic volume index of production cost and dividing the cost ratio by the true volume index of production cost gives the economic price index of production cost. On the other hand, cost ratio decompositions to Salter measures and economic indices connect the partial Salter approach, presented at the beginning of this section, to the general theory of production and costs.

The decompositions (4) and (5) above would be reformulated by using the concepts of true price and volume indices of scale and marginal cost and output ratio. Further they may be decomposed due to factors of technical change only measured at certain prices and an output level and factors due to different output levels and prices; see Karko (1986).

3. EMPIRICAL APPLICATION

3.1. Preliminaries

In this chapter we shall apply the production function framework discussed above to a specific industry namely the Finnish brewing industry. Both the frontier production function and the short-run industry production function are estimated for this industry. This yields an interesting and fruitful possibility to compare the results of the two production functions and the efficiency measures connected to them. Earlier studies have not analysed both the production functions for the same industry.

First we discuss the demands to be placed on the data, describe briefly the data used and the major features of the industry subjected to this analysis. Secondly we discuss the suitability of the data and the industry as the empirical data for this study and, finally, we define the variables used. The industry is described in more detail in Summa (1971), (1972) and (1985) and the data in Appendix 1.

High demands are placed on data intended for the empirical testing of the methods used in this study. (1) Data on clearly defined units on a disaggregated level is needed. (2) The data must be of high quality and accuracy, because errors in measuring the variables easily lead to false conclusions, for example as regards the determination of the distributions of partial productivity or estimation of the frontier production functions.[1] (3) The length of the time series must also be taken into consideration in a study of technical change and structural development.

---

1) Problems caused by "outliers", e.g. due to measurement errors, are discussed in Førsund et al. (1980), Schmidt (1985) and Førsund (1985).

One needs a period of time long enough for a significant proportion of the fixed assets of the industry to be renewed. (4) When evaluating the applicability of methods, it is important to find clearly definable, different phases in the development of the industry, regarding e.g. its production growth rate, investment intensity, technical change, relative input prices and competitive situation. (5) Furthermore, it is of importance that there be a "sufficient" number of units to be observed within the industry, so that more varied analyses and an improved statistical treatment may be achieved.

The data used was gathered from the Finnish brewing industry[1)] and covers the period of 1954-1984. The observation unit is a brewing plant. The basic data was obtained from each plant's annual Industrial Statistics information forms. These data were augmented, e.g. regarding the years 1954-1958, by data obtained from the balance sheets of each firm. Since each plant was, at that time, an independent firm it was possible to augment the data this way. Because of the analysis techniques applied, special care was taken to ensure that the basic data was of high quality. All of the plants were visited and their management interviewed. The main emphasis at this phase was, besides checking and supplementing the data, on the technology applied and investments made.

All plants were included at the preliminary stage, but for the analyses reported later in this study, plants were chosen according to the criteria used by Airaksinen (1977):[2)]

1. The plant must be included in the industrial statistics under the heading "Brewing Industry".

---

1) In 1954-1970 the ISIC sector 2132 and since 1971, sector 31332.

2) Airaksinen (1977), p. 84.

2. The brewery must produce beer with an ethylene content high enough to be covered by the Act on Alcoholic Beverages.

3. The plant must have been operating for a minimum of eight years during the period 1954-1984.

Under these criteria, 18 plants out of the original 29 were chosen.[1)] Data on these plants was available for different lengths of time. The output of the plants included accounted for over 95 % of the output of the entire sector in all the years observed.

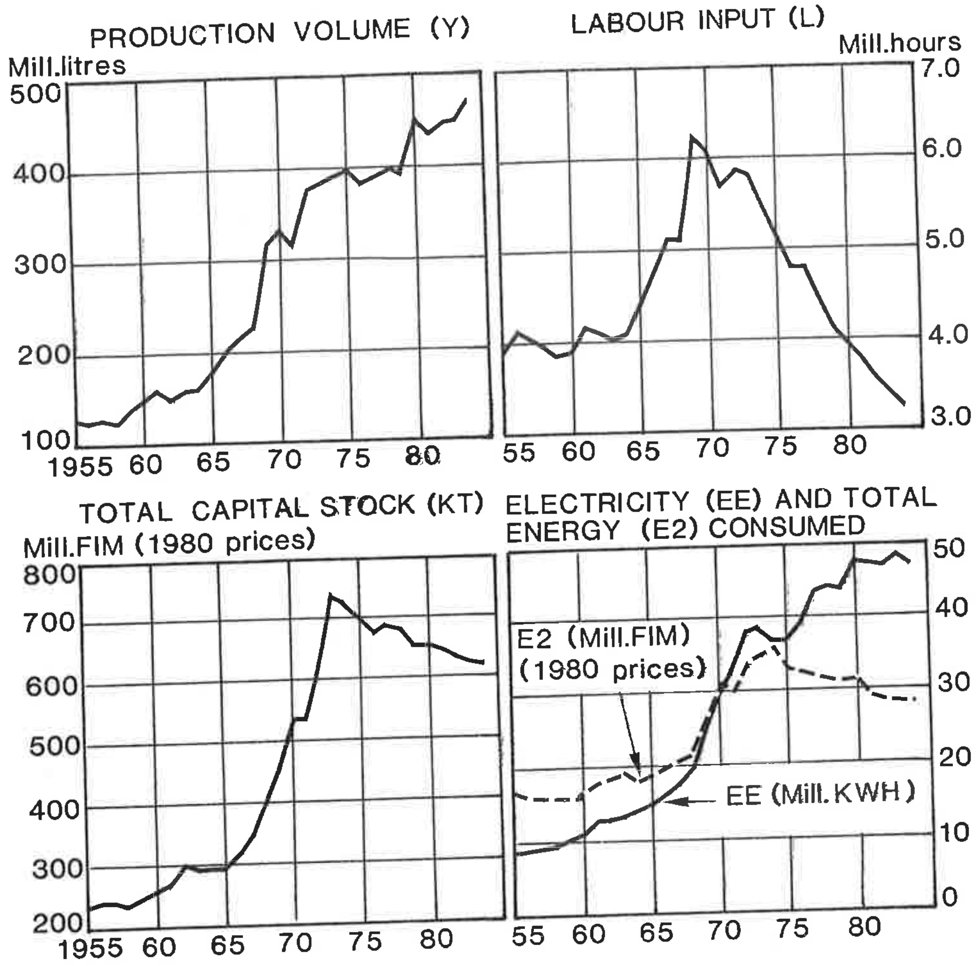
The brewing industry represents a typical process industry. Its input factors and end products are homogeneous in comparison with most other industries and relatively easy to subject to accurate measurement. The dominant final product is beer. However beer is produced to some degree jointly with soft drinks and mineral waters. Some phases of the production process, e.g. bottling, are used for all three products. Since the Industrial statistics yield data for each plant on the whole we have no information to distinguish resources used for beer and resources used for soft drinks and mineral waters.

The technology applied is capital embodied in its nature. After an investment has been made, the substitutability between the various factors of production is very low. The management interviewed considered the development of the industry to follow the lines of a typical putty-clay technology. The latest advances in technology at each point in time were potentially available to all the breweries included.

---

1) The plants not included this study were analysed in Puputti (1983).

Figure 3.1: Development of production volume and labour, capital, electricity and total energy inputs.



In the development of the industry, the observation period can be divided into three distinct phases (see Figure 3.1): the phase of slow, steady growth in 1954-1967, the phase of heavy investments and radical production volume growth in 1968-1972 and the phase consisting of the years 1973-1984 when production growth was slow but steep changes ocurred in input prices (see Table 3.1) and the sharpening competition forced the firms to intensify efficiency improvements. Because the years 1968-1972 were exceptional in many ways, we emphasise in reporting the results based on the total period and on the subperiods of 1955-1967 and 1973-1984.

The new legislation on alcoholic beverages, in force since the beginning of 1969, made a radical change in the availability of medium-strong beer, for which reason the consumption rose to a new, markedly higher, level. The average growth of output during the first two phases, excluding the above-mentioned shift in production level, was of the same magnitude as that of the overall volume growth in the Finnish industries. Fixed assets at constant prices increased slowly in the years 1954-1966, while the years 1967-1972 saw the value of machinery and equipment, at constant prices, more than doubled as the brewery firms prepared for the shift in production level. It is unlikely that there could be any other industrial sector comprising more than ten companies in Finland that would ever have more than doubled its fixed assets within three years.

Table 3.1 Development of relative factor prices[1)]

Year	L/EE	L/E2	L/KT	EE/E2	EE/KT	E2/KT
1955	1.00	1.00	1.00	1.00	1.00	1.00
67	1.91	1.61	2.69	0.84	1.41	1.67
73	2.37	2.65	5.23	1.11	2.20	1.98
84	3.43	2.58	23.25	0.75	6.77	9.03
1973	1.00	1.00	1.00	1.00	1.00	1.00
84	1.45	0.97	4.44	0.67	2.98	4.56

At the beginning of the observation period, the number of plants was 28, in 1972 there were 13 and in 1984 11 plants. The average annual output per plant grew dramatically.[2)] This is an important observa-

1) The factor prices used are defined in Appendix 1.

2) See Summa (1985), p. 13.

tion, because according to several studies, considerable economies of scale can be achieved in the brewery industry.[1] The plants are fairly equal in size (see Figure 3.2)[2], which is likely to diminish the efficiency differences due to economies of scale. The firm concentration ratio deviates only slightly from that of the individual plants in spite of the fact that mergers yield multiplant firms.

Marked changes occurred in the market for the end-products during the observation period.[3] At the beginning each firm had a sales district, appointed by the state-owned monopoly company Oy Alko Ab, allowing them a partial regional monopoly. This system was gradually relaxed and eventually abolished. The sharpened competition has been a major factor in adding pressure towards increased efficiency. From the point of view of a brewery, the price of the end-products could be largely considered exogenous due to the competitive structure of the industry. In fact, the market situation has just the same typical properties as the market socialists Lange and Taylor (1938) and Lerner (1944) in the 1930s connected with perfect competition.

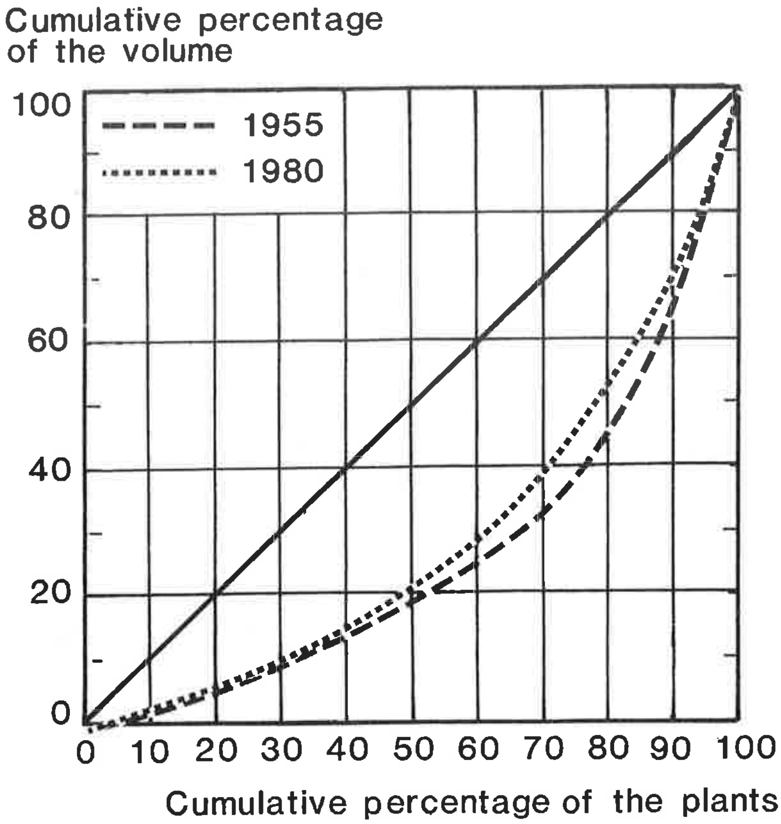
Possibilities of an individual brewery to influence the input markets are rather limited. Each brewery pays nearly the same price, mostly determined by the world market, for its raw materials. The price of labour in an individual firm, in an individual industry even, can be

1) E.g. Ribrant (1970), Gabrielsson (1970), Pratten (1971), Seeringer (1975a) and (1975b), Airaksinen (1977), Schwalbach (1981), (1984a), (1984b) and (1984c).

2) A comparison with Vuori (1981) shows that the distributions is more equal than in the main industrial sectors in Finland in general.

3) The development has been presented in detail in Summa (1971) and Österberg (1974).

Figure 3.2: Lorenz curves according to production capacity in 1955 and 1980.



seen as largely exogenous too, because wage agreements are negotiated at the national level.[1] The costs of investment and financing can also be considered almost identical in each firm within this industry. Regarding the whole industry, large changes occurred in the relative input prices in the course of the observation period; especially the steady and heavy rise in labour costs and the high level shifts in energy prices in the 1970s caused adjustment pressures (Table 3.1).

In conclusion we may note that both the industry under study and the data used fairly well fulfill the requirements mentioned earlier in

---

1) See Lastikka (1977).

this section.[1] Among the many alternative operational variables constructed for the empirical analysis, only the following were chosen for this report: Production volume (Y) is expressed in equivalent beer litres. The soft drink volume was translated into a beer equivalent by weighting it using the price ratio soft drinks/beer for each year. The equivalent litre volumes were added together. Labour input (L) is based on hours worked by all workers. Various other labour input variables were used in analysing the data, but the results were not sensitive to the different specifications of the labour input variable. The total capital input (KT) was constructed by the perpetual inventory method and includes machinery, equipment and buildings.[2] Two different measures were used for energy consumption. The electricity consumption (EE) reflects not only the usage of energy but the degree of sophistication of the technology applied. The total energy consumption (E2) takes into account all the kinds of energy used during the observation period. Fore more details see Appendix 1.

3.2. Description of the structure

In the Salter-Johansen framework there are several ways to describe the intra-industrial efficiency and the changes occurring in it. In practice, the exact objective of the analysis[3] and the available

---

1) See also Summa (1985), pp. 17-18.

2) For a detailed representation of the construction of the capital variable see Summa (1985) pp. 132-149.

3) Teague and Eilon (1973) distinguish four reasons for measuring efficiency at the firm level: (1) strategic, (2) tactical, (3) planning and (4) internal management purposes.

data determine the approach to be chosen[1]. In this section we briefly present various potential ways of illuminating the distribution of productivity and employ partial input coefficient distributions and capacity distributions in describing the development of the brewing industry. The more comprehensive methods, which take into consideration more than two inputs simultaneously, will be taken up in Section 3.3.

In Table 3.2, the alternative ways of measuring intra-industrial efficiency are divided into two main categories: the first is based on quantitative data on input factors and output; the second further presupposes the availability of price variables in addition to their volumes. Partial productivity usually indicates the volume of output per a certain volume of one input factor. Partial productivity, when used as a separate index figure, only gives us information about one factor and its development. Partial productivities are on the other hand functionally connected to the production function: if $y=f(x)$, $y/x_i=\bar{f}(x)$, $(i=1,...,n)$, so the partial productivities depend on the production function. Under cost minimisation partial productivities are also connected with factor prices and output: $y/x_i=X(w,y)$.

The comparison of various partial productivities between the various plants may give us conflicting impressions of the process of change going on in the background. Partial productivities often are used on the assumption that all productivities, other than the one currently under analysis, are identical in all plants. This assumption is rarely the case in the real world.

1) See e.g. Wohlin (1970), Todd (1971) and (1985) and Särskilda Näringspolitiska Delegationen (1979) on the choices made on the basis of various objectives and data.

Table 3.2: THE FRAMEWORK IN ANALYSING THE INTRA-INDUSTRIAL EFFICIENCY BY USING MICRO DATA

USAGE OF INPUT AND OUTPUT VOLUME DATA

VOLUME OF INPUT		Input productivities	VOLUME OF OUTPUT
(1) Raw materials (R)	→	$\frac{R}{Y}$	
(2) Labour (L)	→	$\frac{L}{Y}$	← Y
(3) Energy (E)	→	$\frac{E}{Y}$	
(4) Capital (K)	→	$\frac{K}{Y}$	
(5) Other production factors (A)	→	$\frac{A}{Y}$	
(6) Total factor usage (L,K,R,E,A)	→	Total factor productivity index	← Y
(7) Production function and its residuals	→	F (x) · u	

USAGE OF COST DATA

COST OF INPUT USAGE		Input shares	VALUE OF OUTPUT
(8) Raw materials (RW_R)	→	$\frac{RW_R}{YP}$	
(9) Labour (LW_L)	→	$\frac{LW_L}{YP}$	← YP
(10) Energy (EW_E)	→	$\frac{EW_E}{YP}$	
(11) Capital KW_K)	→	$\frac{KW_K}{YP}$	
(12) Other production factors (AW_A)	→	$\frac{AW_A}{YP}$	
(13) Total cost ($LW_L+KW_K+RW_R+$ $EW_E + AW_A$ = 1 - (TC/YP)	→	Total cost efficiency	← YP
(14) Cost and profit functions and their residuals	→	$C(y,w)\cdot u$ $\Pi(p,w)\cdot u'$	

Depending on the point of view chosen, it may be advisable, when analysing the efficiency structure of an industry, to group the factors of production according to their degree of flexibility in the short-run.[1] Raw materials and labour are often considered to be variables. A firm may excercise only partial influence on its energy input in the short-run, because this often is largely determined by the technology applied. This variable may be termed semi-fixed. Capital is, in the short run, a fixed production factor.

Partial productivities (1)-(5) in Table 3.2 may be analysed individually or in various combinations. However, we often are interested in the total efficiency of several important inputs. This combined effect may be put in a compact form by constructing total factor productivity indices[2] or by using the production function as below in Section 3.3. The production function may be derived from economic data when the decision making is based on profit maximisation, cost minimisation or some other clearly defined objective.

When the prices of the inputs are available, in addition to their volumes, we can describe the efficiency structure by cost distributions, factor by factor, or by summing up the relevant coefficients. Shephard's (1953) duality theory shows that the information contained in both cost and production functions is equivalent if the firm is

---

1) For more detailed analysis, see Lehtonen (1976) and Maltzan (1978).

2) See e.g. Nadiri's (1970) and Diewert's (1982) survey articles on different approaches to the theory and measurement of total factor productivity, Simula (1983) for an application to the Finnish forest industries and Fischer (1984) for an application to the German chemical industry.

minimising costs. This result serves as basis for a widening theoretical and empirical analysis, even in the study of efficiency distributions.[1)]

Salter uses a histogram to illustrate productivity distributions of an industry. On the ordinate axis he introduces one input coefficient at a time over the plants and on the abscissa axis he has relative or absolute levels of production. In this distribution, called the Salter diagram, the plants are generally presented in the order of increasing values of the input coefficients. Each plant has its own retangle in the histogram giving us its share in the total production of the industry and its input coefficient. The information on the abscissa axis remains generally unchanged in a certain cross section analysis, but the shape of the distribution on the ordinate axis varies with the variations of each individual input coefficient analysed. A histogram with average costs of the plants on the ordinate axis instead of input coefficients is called a Heckscher diagram.

In spite of its simplicity, the Salter diagram provides a basis for a variety of empirical analyses. The management of each plant is interested in the plant's "ranking" in the distribution. An industrial policy maker may find important information at both ends of the distribution. The deviations expressed by the distribution often reveal potential pressures towards change in the industry. The form of the

---

1) See Fuss and McFadden (1978).

distribution in an industry undergoing rapid production growth may be quite different from that of declining industries.[1)]

The intensity of technological progress in each coefficient under observation is illustrated by the difference in efficiency between the latest best-practice plant to enter the industry and the older ones. The distribution also gives information on the connection between productivity and plant size i.e. whether small/large plants are concentrated at the same end of the distribution.

Further aspects, important from the point of view of empirical analysis in each individual case, may be added to the histograms, such as some aspect of technology (e.g. the process applied), the owner of the plant in the case of multiplant firms, the vintage of the plant and geographical area.[2)] Analysis of cross sections of successive years illustrates the dynamics of the industry. Statistical parameters may also be used in analysing the distributions.

The examples given above on the application of the Salter distributions show how well they are suited for illustrating the intra-industry dynamics as well as organising and understanding the data and even for testing simple behavioural hypotheses. The gravest limitation here is the partial character of the Salter point of view: one analyses the productivities of the individual inputs separately, not the total impact of all inputs, the latter would be of vital importance from the viewpoint of the firms'

---

1) See Salter (1960), p. 82.

2) See Summa et.al. (1985).

goal function. It is our intention here to proceed from the partial input coefficient distributions, via capacity distributions, to the measurement of productive efficiency using production functions.

Among the various alternatives in Table 3.2, we in this report concentrate on the efficiency distributions based on input and output volume data. In practice all the plants pay nearly the same price for their inputs and, it is, therefore, justifiable to use the volume variables. We have already noted above the close connection between the production and cost functions via the duality relationships.

Among the partial productivities labour and energy are given the closest analysis as examples of variable inputs while capital is an example of a fixed input. Together with raw materials, these inputs account for the highest expenditures. There were distinctly smaller differences in the productivity of raw material input between the plants than in the labour, energy and capital inputs. Even though the cost share of the raw materials is high, the input is not a vital productivity factor in the industry under observation[1)] and since there is practically no ways to substitute raw materials for other inputs, raw materials were not included in the production function analyses.

Among the many factors, the productivity of <u>labour</u> has traditionally been studied intensively. This interest does not entirely depend on the large share of labour costs in the total production costs. An ap-

---

1) In many industries the productivity in the usage of raw material input may be strategically a most decisive factor, see e.g. Simula (1983) pp. 41-42.

proach often resorted to in competitiveness analysis is that the firms buy their raw materials at an identical (world) market price and can only have a slight influence on the productivity of raw material inputs because, for technical reasons, the input-output relationships are fixed and nearly identical in all the competing plants. The wage and labour skill levels, on the other hand, vary heavily from country to country and, furthermore, management is assumed to be able to exert some influence on the productivity of labour. The continued efforts to increase the productivity of labour have been partly the result of the increase in the relative price of labour compared to that of the other inputs (see Table 3.1).

Even empirically significant differences in the productivity of labour have been found ex post between plants which ex ante, in terms of technology chosen, have been practically identical.[1] These differences have been analysed and explained by reference to X-inefficiency.[2] Differences in competitiveness between various firms and especially plants in the same industry across countries are often explained by reference to considerable differences in the productivity of labour (Pratten 1976, Panic 1976). The focus on labour is further explained by the availability of detailed and reliable data on labour inputs, in comparison to other inputs.

The Salter diagram of labour input coefficient L/Y is illustrated in Figure 3.3 for the cross sections of 1955, 1967, 1973 and 1980. The labour productivity is the inverted value of the input coefficient.

1) Cf. Rodas and Humberg (1980).

2) See Leibenstein (1966) and (1975) and Shen (1984).

Productivity has improved throughout the period under observation. The embodied and disembodied effects on savings in labour input cannot be distinguished directly from the observations. However, since the plants did not make any major long-range investments in 1955-1967 the 'Horndal effect' should be important. After 1967 the distinct shift and increase in the productivity level in connection with the large investments made in 1968-1970 supports the view that this was mainly an embodied effect, typical of process industries. The development in the 1970s is probably more a mixture of the embodied and disembodied productivity increase.

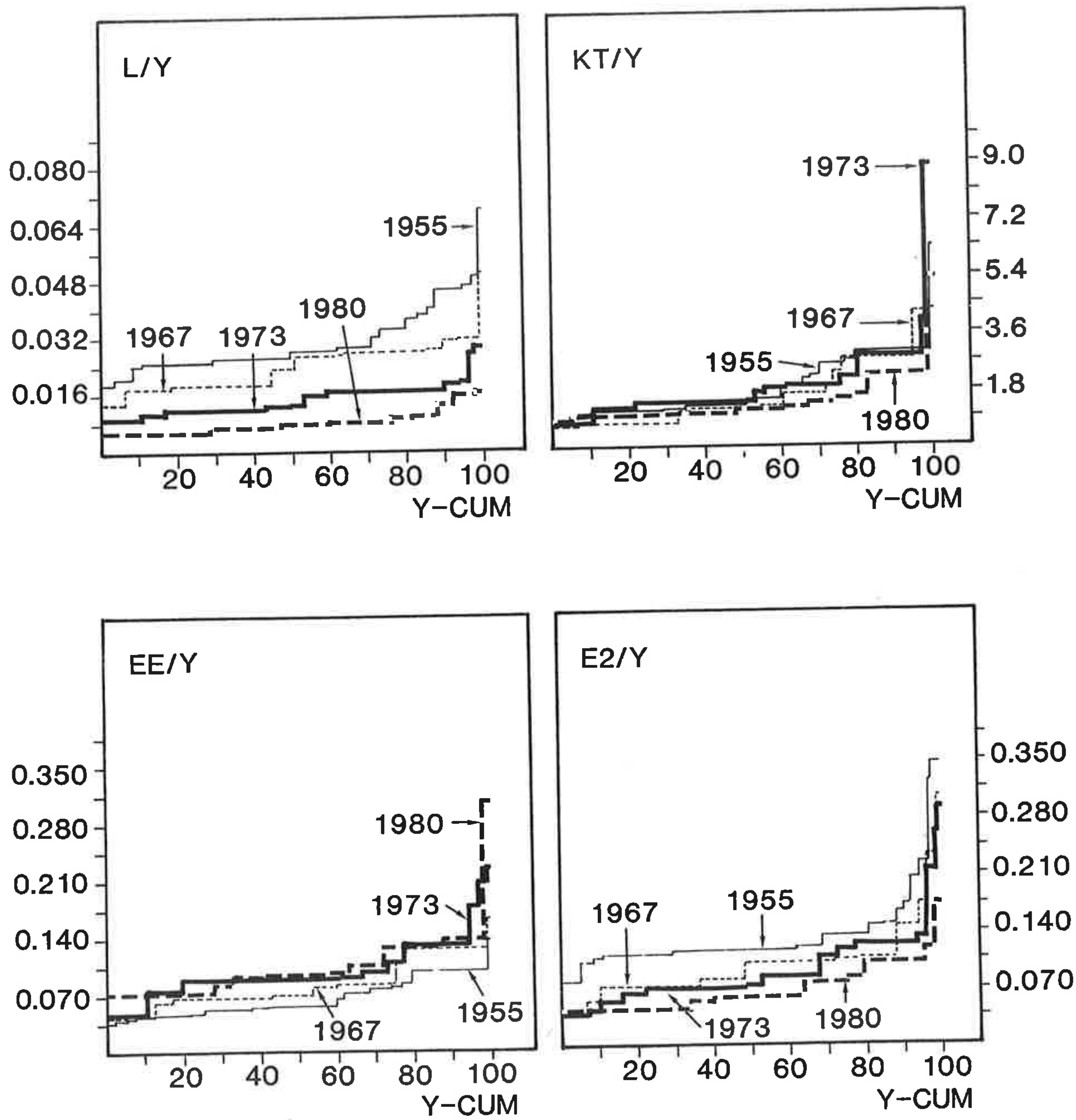
The productivity differences between the plants were large in the first period but clearly diminishing later during the observation period. Small units are in general, throughout the observation period, concentrated at the low productivity end of the distribution. Large units, as a rule, are of a later vintage, which means that they are using the latest technology and benefiting from the effects of scale, increased mechanisation and process automation.

Energy intensiveness is, to a fairly high degree, linked with the technology determined by the investments made. In the short run the amounts of energy used are partly fixed (heating, lighting etc.) and partly dependent on output volume. Energy may thus be considered a semifixed production factor. In a process industry, energy intensity often is also a good indicator of the type and vintage of the technology applied.[1] We employ two different variables to measure energy consumption, electricity (EE) and total energy used (E2), which illuminate different aspects of technological progress.

---

1) E.g. Summa et.al. (1985).

Figure 3.3: Salter diagrams of labour, electricity, total energy and capital input coefficients in 1955, 1967, 1973 and 1980.



In empirical studies in the field of production theory, electricity is often used as a proxy for the total energy consumption or for the capital variable. The energy input can often be measured precisely and is assumed to gain wider application as technology becomes more capital intensive. In our case the substitution pressure, caused by changing relative prices (Table 3.1), was found to be a factor contributing to the growth of the capital input; the increase in the use of electrical power is, of course, an integral part of the same development.

The shape of the Salter diagram for total energy remained almost unchanged in the comparison years (Figure 3.3). In some intervals electricity consumption increased slightly faster than the output, i.e. the electricity intensity of production increased. This trend is a natural one, because the relative price of energy, in comparison to other factors of production, sank during the observation period and attempts to save labour increased electricity intensity. There seem to be no clear scale economies in energy use. Small units appear at both ends of the distribution.

Until the dramatic increase in the absolute and relative price of energy in the year 1974, the electricity consumption worked well as proxy for the <u>total energy use</u>; the electricity intensity increased only slightly faster (Figure 3.1). Since then the intensity of electricity consumption has remained nearly constant in contrast to the sharp decrease in other types of energy. While the relative price of electricity fell during the period of observation the relative price of oil and coal increased strongly after 1973 inducing energy saving programs and investments in several plants. The result is a very clear decrease in total energy use between the years 1973 and 1980 (Figure 3.3). The same trend continued in the early 1980's.

Differences in <u>capital</u> productivity are of primary importance for the analysis of intra-industry structure. According to the traditional neoclassical production theory, capital consists of homogeneous technical units which must be combined with other factors of production to create output. In practice, however, capital goods are heterogenous, designed for a specific purpose. To alleviate this conflict between production theory and reality, Johansen and Salter introduced their putty-clay framework.[1)]

Despite its vital importance, there are few studies on the productivity of capital in the economic literature. A partial explanation for the scarcity of empirical research - especially using micro-data - in production theory may be found in the fact that it is difficult to define and measure the capital variable and that data is rarely available. In the present study, the analysis of the productivity distribution of the capital variable is of particular interest, not only because of the putty-clay framework, but because the observation period includes both a rather static phase 1954-1967, a dynamic phase characterised by significant modernisation and expansion of the capital stock, 1968-1972 and the years 1973-84 when increasing the productivity of the machinery was a major target in all plants.

Technological progress in the Finnish brewing industry has been steady, without major innovations or breakthroughs. All plants have potential access to the latest technology, though in practice, only at that point in time when investments are made. A plant is very rarely of a "pure" vintage, because in practice plants, with a couple of exceptions, have been expanded and changed gradually.

---

1) In principle the same difficulty lies on all other input aggregates e.g. labour as noted by Diewert (1980).

As a rule, ex post substitutability is very limited. A large single investment is, in practice, followed by a stream of smaller investments, with the aim of improving the plant's productivity. The character of further investments is determined by substitution pressures and by the new technical solutions for these. Because the price of labour has been rising constantly and faster than the prices of other production factors, efforts to reduce the use of this input comprise the dominant features of the industry studied.[1] New machinery and equipment comes with a higher degree of automation and mechanisation, especially in the more labour intensive process phases, such as the bottling lines.

The replacement of labour by machines has slowed down the decrease of the capital input coefficient. Shortening of the working week has had a similar effect. As the working week shortens, the intensity of use of existing capital may decrease and the stock figures overestimate the input of capital services.[2] As we see in Figure 3.3, the differences in capital productivity between individual plants remained fairly large.

A capacity distribution diagram[3] combines information from two Salter diagrams (Figure 3.4). The input coefficients are measured along the axes. The size of each square is proportional to the capacity of the corresponding plant. The range of variation for both input coefficients is shown simultaneously. It is of interest to see, for

1) The analogous development applies to the energy input since 1974.

2) See Solow (1957).

3) Sato (1975) calls this diagram the efficiency distribution.

instance, whether the capacity is in a south-west/north-east direction or in a south-east/north-west direction. Putting the observations of two years into the same diagram allows us to look at changes in structure between two different points in time. In Figure 3.4 empty squares illustrate the capacity distribution in 1980 and the black ones represent the situation at the beginning of the observation period, in 1955.

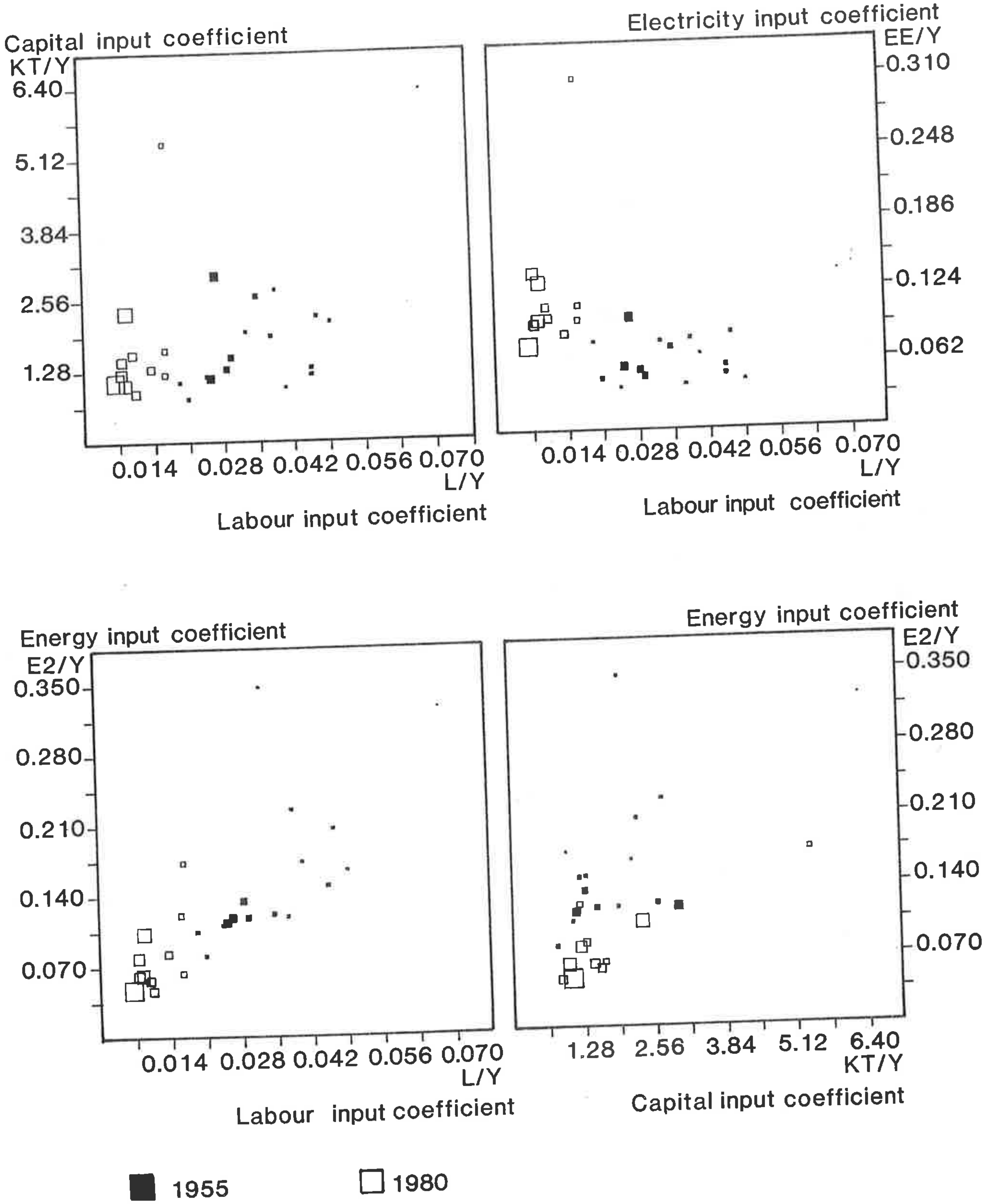
It is possible to illuminate different development processes by using two-dimensional diagrams. For instance, neutral technical progress or increased exploitation of economies of scale may be expressed as the simultaneous reduction of both input coefficients. On the other hand, if structural change has been characterised by a transformation of the structure in the north-west/south-east direction, there should be a substitution process between at least two factors behind this structural change due to:

1 Development of the ex ante or choice of technique production function or

2 development of relative factor prices influencing scrapping and choice of technology in new equipment

The capacity distributions in 1955 and 1980 are shown in Figure 3.4 for different pairs of inputs. The capacity distribution has moved considerably between 1955 and 1980, especially in the labour and total energy saving direction. The average value of the input coefficient for labour, for the industry as a whole, has, from 1955 to 1980, decreased by 71 per cent and 47 per cent for energy. The shape of the distribution has changed due to the presence of large capacity in large units in 1980 with a corresponding high labour productivity. The largest unit in 1980 is also the most efficient.

Figure 3.4: Capacity distributions in the input coefficient space in 1955 and 1980.



In the labour-capital dimension (Figure 3.4) there is a parallel movement of the structure towards the capital axis and in the capital-total energy dimension there is both a movement towards the capital axis and a movement towards the the origin indicating a mixed process of substitution, technical change and scale factors. Finally between total energy and electricity a typical substitution process has taken place.

3.3. Frontier production functions

3.3.1. Deterministic approach

3.3.1.1. Estimation procedure

The main objective of this study is to analyse the development of intra-industrial structure and the character of technological progress using a putty-clay framework. The approach chosen requires analytical methods which might be different from those normally used in inter-industry comparisons of structural and technological change.[1] In this section we use a frontier production function as the main tool of analysis. Earlier, in Chapter 2, the theoretical motivation for the frontier production function was treated at a general level, while here we present a specific application. From the large variety of estimation methods, treated briefly in Section 2.3.2., we find the deterministic approach most suitable in our case.[2] We thus "estimate" the frontier by using a programming method solving an LP-problem with on-or-below-frontier constraints. The method was first suggested by

---

1) Cf. Carlsson (1972).

2) The advantages of using the deterministic instead of the stochastic model is discussed in Førsund (1985).

Aigner and Chu (1968) for Cobb-Douglas production functions but later generalised by Førsund and Hjalmarsson (1979) to homothetic production functions. Thus the method chosen for this study allows a neutrally variable returns to scale[1], a prerequisite for an analysis of the development of optimal scale, an important aspect of our analysis. For purposes of comparison, however, we also present estimates of a stochastic frontier.

The LP computational framework can be maintained even for general functional forms, such as the translog:

$$\ln f(x^j) = \ln a_0 + \sum_{i=1}^{n} a_i^j \ln x_i^j + \frac{1}{2} \sum_{i=1}^{n} \sum_{k}^{n} \gamma_{ik}^j \ln x_i^j \ln x_k^j$$

where we index plants by $j=1,\ldots,m$ and inputs by $i=1,\ldots,n$. The translog functional form is a second order (local) approximation for any production function.[2] Its scale and substitution elasticities are variable, as opposed to the Cobb-Douglas and CES-functions, for which they are constants. Note that the Cobb-Douglas production function is a special case of a general translog function. On the other hand, CES-functions are second and Cobb-Douglas first order approximations to a translog function. The usefulness of the approximation character of the translog form might have been exaggregated just for the local nature of the approximation. Numerous Monte-Carlo studies show that

1) See Førsund and Hjalmarsson (1979).

2) See Lau (1974), Lehtonen (1976) and Diewert (1980).

the approximation may be poor if the true elasticities of substitution deviate substantially from one. However, these shortcomings have not diminished the popularity of the translog function in either the estimation of average production or that of cost functions.[1] The technical difficulties related to the general functional forms and our need to emphasise the scale properties have influenced our choice of a homothetic function with a Cobb-Douglas kernel which allows for variable elasticity of scale. We have as our starting point a frontier function, prespecified to be a homothetic function of the general form

$$F(y) = f(x)\cdot u \qquad (0 \leq u \leq 1) \tag{1}$$

where y is the rate of output, x is the vector of inputs, F(y) is a monotonically increasing function, f(x) is a homogeneous function of degree 1 and u is a stochastic variable implying input-neutral differences between units with respect to what they can achieve with their inputs.

To tackle the importance of the economies of scale we specify the transformation function in the following form[2]

$$\ln F(y) = \alpha \ln y + \beta y \tag{2}$$

As regards the kernel function, several tractable production functions may be employed. As the kernel function $f(\cdot)$ we use a linear homogeneous Cobb-Douglas function. A homothetic function with a Cobb-Douglas kernel

1) See Guilkey and Knox-Lovell (1980) and Thursby and Knox-Lovell (1978).

2) Cf. Zellner and Revankar (1969).

function is a suitable, simple functional form for practical calculations even in analyses of technical change as has been shown by Førsund and Hjalmarsson (1979). It serves our purposes well because a homothetic function permits a variable elasticity of scale which depends on output, cf. Førsund (1974).

According to our chosen approach, the observations should be close to the frontier. We, therefore, seek to minimise the simple sum of deviations from the frontier with respect to input utilisation after logarithmic transformation, subject to on-or-below-frontier constraints.

For computational convenience the following increasing function in the efficiency measures u_i is to be maximised:

$$\sum_j (\alpha \ln y^j + \beta y^j - \ln A - \sum_i a_i \ln x_i^j) \tag{3}$$

subject to the on-or-below-the-frontier constraints:

$$\alpha \ln y^j + \beta y^j - \ln A - \sum_i a_i \ln x_i^j \leq 0 \quad j=1,\dots,m \tag{4}$$

and

$$\sum_i a_i = 1 \qquad \text{(the homogeneity constraint).}$$

With this specification the estimation problem is reduced to the simple problem of solving a standard linear programming problem.

In this study technical progress at the frontier is captured by introducing trends in all the parameters of the frontier function. Changes in the optimal scale over time are according to previous studies especially interesting.[1] Using trends in both the scale and the kernel function parameters we should be able to illustrate the character of technical progress. Using L for labour, K for capital, E for energy, A for the constant term, a_i for the kernel elasticities and the specification given above, we get the following frontier function:

$$y^{\alpha-\gamma_5 t} e^{(\beta-\gamma_6 t)y} = Ae^{\gamma_4 t} L^{(a_1-\gamma_1 t)} K^{(a_2-\gamma_2 t)} E^{(a_3-\gamma_3 t)} \quad (6)$$

where α and β are the scale function parameters and the returns to scale properties are given by the elasticity of scale function[2]

$$\varepsilon(y,t) = \frac{F(y,t)}{y.F'(y,t)} = \frac{1}{\alpha-\gamma_5 t + (\beta-\gamma_6 t).y} \quad (7)$$

which yields the following expression for the technically optimal scale, $\hat{y}$:

$$\hat{y} = \frac{1-\alpha+\gamma_5 t}{\beta-\gamma_6 t}$$

Using the functional form given above, originally proposed by Zellner and Revankar (1969), and without introducing a specific distribution

---

1) See e.g. Pratten (1971).

2) For the derivation of the technically optimal scale in inhomogeneous production functions see Førsund (1974).

for u, the computation of the frontier function reduceds to the solution of a simple linear programming problem. The objective function to be minimised is

$$\sum_{t=1}^{T}\sum_{j=1}^{m} (\ln A) + \gamma_4 t + \sum_{i=1}^{3}(a_i-\gamma_i t)x_i^j(t) - (\alpha-\gamma_5 t)\ln y^j(t) - (\beta-\gamma_6 t)\cdot y^j(t) \tag{8}$$

The following are the constraints in our linear programming model[1]:

First, since all observations must be on or below the frontier, we get from (8), T·n constraints

$$\ln A + \gamma_4 t + \sum_{i=1}^{3}(a_i-\gamma_i t)x_i^j(t) - (\alpha-\gamma_5 t)\ln y^j(t) - (\beta-\gamma_6 t)\cdot y^j(t) \geq 0 \tag{9}$$

Secondly, the homogeneity constraint gives

$$\sum_i a_{i,t} = \sum_{i=1}^{3}(a_i - \gamma_i \cdot t) = 1 \quad (t=1,\ldots,T) \tag{10}$$

this is due to the linear homogeneity of the kernel function.

Thirdly, if the homogeneity constraint holds for all periods Equation (10) implies

$$\sum_{i=1}^{3} \gamma_i = 0 \tag{11}$$

1) Cf. Førsund and Hjalmarsson (1979) where the two factor case is introduced.

Fourthly, we restrict the kernel elasticities with the trends $a_{i,t}$ to the interval [0,1]. Using (10) and (11) we get the constraints

$$a_i - \gamma_i T \geq 0 \qquad i = 1,2,3 \tag{12}$$

Fifthly, the scale parameters with the trends should be non-negative

$$\alpha - \gamma_5 T \geq 0$$

and (13)

$$\beta - \gamma_6 T \geq 0.$$

Sixthly, it is reasonable to assume that the production function has classical properties[1], which imply an S-shaped production surface and non-negative marginal productivities, yielding the following parameter restrictions

$$\alpha, \beta, a_1, a_2, a_3, \gamma_4, \gamma_5, \gamma_6 \geq 0.$$

If α and β are non-negative the production function follows a regular ultra passum law (in the weak sense) which means that the scale elasticity, if it varies, is decreasing from values larger than one through one to values smaller than one along an expansion path yielding a U-shaped average cost curve.[2] If α or β is negative there is no well-behaved production function since the elasticity of scale is discontinuous.

Seventhly, γ_1, γ_2, and γ_3 and ln A are unrestricted.

---

1) Cf. e.g. Frisch (1965), Intriligator (1971) and Oebe (1974).

2) See Frisch (1965) and Førsund (1974).

3.3.1.2. Main estimation results

The frontier production functions were estimated using the procedure described in Section 3.3.1.1. In the following we present and interpret the main results deriving from alternative specifications of time trends, the energy variable and the time periods. Other aspects regarding the sensitivity of the results will be discussed in more detail in the next Section and the nature of technical change in Section 3.3.1.4.

The model has been estimated, besides the whole observation period 1955-1984, even by subperiods 1955-1967 and 1973-1984. It is of particular interest to study the effects of the exceptionally strong investment boom of the late 1960s and early 1970s and the rapid changes in relative input prices, especially between energy and other inputs occurring since 1974, on technical change. In this analysis we concentrate on the results arrived at by using the labour input (L), total energy consumption (E2) and total capital stock (KT), even though we even studied the effects of other variable choices on the results. The empirical results are presented in Tables 3.3 and 3.4. We name the model with trends in all parameters as the main case (Case 1 and Table 3.4).

Energy variable

Special attention was paid to the definition of the energy input, because the importance of energy, along with the rapid rise in its price, as a factor affecting the competitiveness of a firm grew essentially during the latter subperiod. Another decisive factor was the determined effort to raise the productivity of labour through increased automation and mechanisation, the price of labour rising constantly in relation to that of electricity (EE) and capital (see Table 3.1). Summa (1985) used electricity consumption as proxy for total energy consumption, which usage was largely well-founded when the estimation period did not go beyond 1972. As the situation has changed, an aggregated variable of total energy consumption was used in this study and its behaviour differed radically from the electricity consumption from 1974 on (see Figure 3.1). It is to be assumed that electricity consumption reflects more the nature of technology than the efficiency of energy input usage. The impact of the choice of the energy variable on the frontier production function results is illustrated in Table 3.3.

We compare results for two different energy inputs, electricity (EE) and total energy (E2) respectively. Table 3.3 summarises the results of the two energy variables EE and E2 in the main case. As should be expected the main differences between the two lie in the differing values of the kernel elasticity of energy. Due to the homogeneity restriction the other kernel elasticities must adjust correspondingly. For the two subperiods the capital elasticity is fairly constant while the labour elasticity differs. Taken for the entire period, opposite results are obtained, however.

The scale properties are not influenced very much during the entire period and the first subperiod. During the last subperiod 1973-84, however, there is a large difference in optimal scale between the two cases.

Below we will concentrate on total energy as the energy input. It is more reasonable to apply a Cobb-Douglas kernel function with total energy than with electricity as the energy variable since the degree of substitution

Table 3.3: Estimates of the frontier production function using electricity (EE) and total energy (E2) as energy variable. Main case.

$$Y^{\alpha-\gamma_5 t} e^{(\beta - \gamma_6 t)y} = Ae^{\gamma_4 t} L^{(a_1 - \gamma_1 t)} KT^{(a_2 - \gamma_2 t)} E^{(a_3 - \gamma_3 t)}$$

	1955-1967		1973-1984		1955-1984	
	EE	E2	EE	E2	EE	E2
ln A	-2.51	-4.19	-1.03	-1.27	-1.05	-1.98
Trend A	0.0	0.07	0.0	0.0	0.0	0.0
a_1	0.29	0.49	0.94	0.24	0.47	0.46
Trend a_1	0.01	-0.02	0.01	0.01	0.01	0.01
a_3	0.39	0.11	0.0	0.73	0.22	0.07
Trend a_3	-0.02	0.00	-0.00	-0.01	-0.00	-0.01
a_2	0.32	0.40	0.06	0.03	0.31	0.47
Trend a_2	0.01	0.02	-0.01	0.00	0.01	0.01
α	0.37	0.20	0.45	0.54	0.56	0.51
Trend α	0.0	0.0	0.0	0.0	0.00	0.0
$\beta \cdot 10^4$	0.67	0.84	0.01	0.06	0.15	0.14
Trend $\beta \cdot 10^4$	0.03	0.03	0.0	0.0	0.03	0.00
Optimal scale in m^3	9700-18700	9900-18500	51500	72000	29300-65000	36400-71700

between electricity and capital should be small in our case. (A Cobb-Douglas function implies that the elasticity of substitution is equal to one between all inputs). Let us therefore concentrate on the results presented in Table 3.4 in addition to E2-results in Table 3.3.

Trend specification

In Table 3.4 we concentrate on total energy, E2, as the energy input but consider different specifications. Case 1 is the main one and here we allow for changes in optimal scale and biased technical change by specifying trends in all parameters. Thus the impact of technical change is simulated by assuming that the parameters of the frontier production function are time-dependent functions: technical change is characterised by successive gradual changes in parameters shifting the frontier function in a non-neutral way.

In Case 2 it is assumed that technical change is extended Hicks-neutral, i.e. the expansion path of the frontier is left untouched and only the numbering of the frontier function isoquants may change.[1)] Hicks-neutrality is easily specified in our frontier function by setting a trend term in the constant term only. Compared with Case 1, Case 2 allows "testing" for the sensitivity of the trend term specification as a nested hypothesis adopting in principle the same approach as in statistical hypothesis testing.

The main case shows a fairly weak trend in all kernel elasticity parameters for all estimation periods (Table 3.3 and 3.4). Over the whole period (1955-1984), technical change is characterised by a decreasing kernel elasticity of labour and capital and a correspondingly increasing energy elasticity. With constant factor prices, this implies that the units should increase the ratio between energy use and that of the other

---

1) See Blackorby, Lovell and Thursby (1976).

Table 3.4: Sensitivity of the frontier production function estimates

$$Y^{\alpha-\gamma_5 t} e^{(\beta-\gamma_6 t)y} = Ae^{\gamma_4 t} L^{(a_1-\gamma_1 t)} KT^{(a_2-\gamma_2 t)} E2^{(a_3-\gamma_3 t)}$$

Case	Kernel function parameters											Scale function parameters					
	Constant term		Labour elasticity			Capital elasticity			Energy elasticity							Optimal scale in m^3	
	ln A	Trend A γ_4	$a_1-\gamma_1 t$		Trend L γ_1	$a_2-\gamma_2 t$		Trend KT γ_2	$a_3-\gamma_3 t$		Trend E2 γ_3	α	Trend γ_5	$\beta.10^4$	Trend $\gamma_6 \cdot 10^4$		
			1955	1967		1955	1967		1955	1967						1955	1967
1	-4.19	0.07	0.51	0.71	-0.02	0.39	0.20	0.02	0.11	0.09	0.00	0.20	0	0.84	0.03	9 900	18 500
2	-1.93	0	0.72	0.72	0	0.25	0.25	0	0.03	0.03	0	0.41	0	0.32	0	18 400	18 400
3	-2.78	0	0.75	0.66	0.01	0	0	0	0.25	0.16	-0.01	0.23	0	0.36	0	21 200	21 200
			1973	1984		1973	1984		1973	1984						1973	1984
1	-1.27	0	0.23	0.16	0.01	0.03	0.02	0.00	0.74	0.83	-0.01	0.54	0	0.06	0	72 000	72 000
2	-1.69	0.01	0.18	0.18	0	0.07	0.07	0	0.75	0.75	0	0.51	0	0.07	0	70 200	70 200
4	-4.48	0.04	0.21	0.40	-0.02	0.23	0.12	0.01	0.56	0.48	0.01	0.26	0	0.17	0.01	45 100	86 400
			1955	1984		1955	1984		1955	1984						1955	1984
1	-1.98	0	0.45	0.28	0.01	0.47	0.24	0.01	0.08	0.48	-0.01	0.51	0	0.14	0.00	36 400	71 700
2	-1.34	0	0.54		0	0.42		0	0.04		0	0.56	0	0.08	0	57 000	57 000

Case 1 = Time series-cross section with trends (<u>Main model</u>)
Case 2 = Hicks-neutral
Case 3 = Main model where plant 17 was deleted for all the years
Case 4 = Main model where plants 2, 18 and 7 were deleted for all the years

inputs. Technical change may, in this sense, be characterised as labour and capital saving and energy using. For an increasing kernel elasticity for labour is noted for the first period (1955-1967) and for energy for the latter period (1973-1984). see Table 3.4. The energy elasticity in the period 1955-1967 is low compared to its value in the period of high energy prices. After the investment boom both labour and capital seem to be relatively more abundant than energy.

The estimated trends in the scale function parameters double optimal scale between 1955 and 1984, from about 36 000 m^3 in 1955 to about 72 000 m^3 in 1984. When using the estimates for the subperiod 1955-1967, it turns out that the optimal scale is in 1967 about twice as large as in 1955. During the last subperiod, however, there is no change in the optimal scale which amounts to about 72 000 m^3. Comparing this with the results for the entire period we find that the optimal scale level is very low during the first subperiod, increasing from about 10 000 m^3 to about 18 500 m^3. On the other hand, the optimal scale level between 1973 and 1984 is almost exactly the same as that of the last year of the entire period, i.e. about 72 000 m^3. It might be interesting to note that observed average output increased from about 7 600 m^3 in 1955 to about 16 700 m^3 in 1967 and from about 32 000 m^3 in 1973 to about 43 100 m^3 in 1984. Thus there seems to be a dramatic change in the optimal scale level due to the investment boom between the two subperiods. The estimated optimal scale levels seem reasonable. In 1984 the largest plant had an output of about 103 000 m^3 and the next largest about 73 000 m^3. It is also interesting to note that the estimated function for the entire period does not seem to represent the technology of the first subperiod very well, on the other hand its results are quite similar to those of the last subperiod. This holds both for scale and, particularly, substitution properties.

The development of the production function can be illustrated as in Figures 3.5-3.7 by drawing the surface of the production function along the average factor ray for the entire period 1955 to 1984) from the origin. This is also clearly illustrated in Figure 3.8, where the elasticity of scale function is plotted along the average rays for the first and last years of all the estimation periods.

Figures 3.5-3.7 illustrates the combined effect of changing marginal elasticities and scale function parameters on the development of the production surface along the chosen, average, factor ray.

In Figure 3.5 it is clearly seen the relative similarity between the production function for the last subperiod and that for the entire period in comparison with the production function for the first subperiod. Due to the different phases in the development of this industry it is not surprising that one production function for the whole period is not able to catch the development for this whole period. It should be said, however, that it explains the scale properties of the last period fairly well.

The surfaces of the production function for the entire period intersect at a fairly high output level. In the range below the intersection point, the production surface in fact moves downwards, but changes its shape corresponding to the increase in the optimal scale.

This movement downwards of the production surface in the low output range reflects the fact that there were no efficient small modern plants in this output range during the last years so data contains no

Figure 3.5: The change in the frontier production function through time. The production function cut with a vertical plane through the origin along the average factor ray. The graphs are based on the estimation of the whole time period and the subperiods 1955-1967 and 1973-1984.

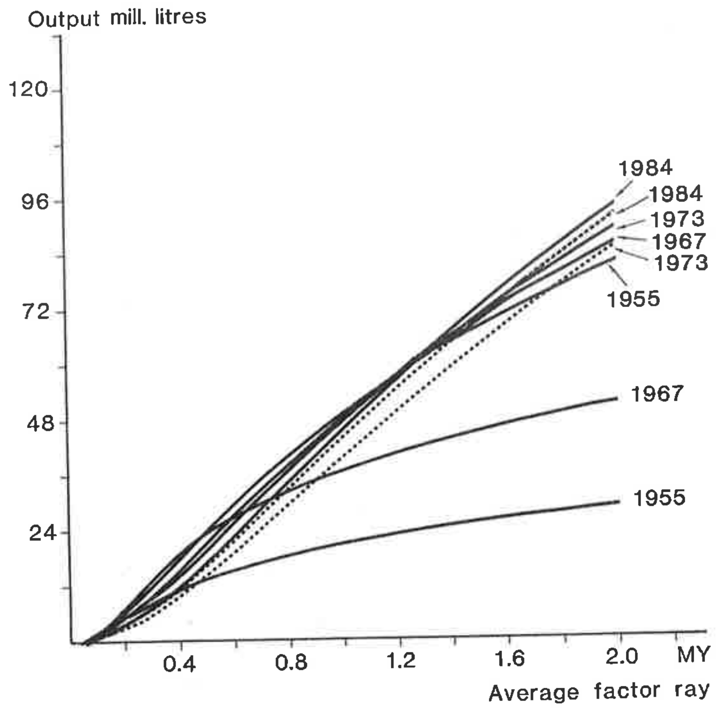


Figure 3.6: The change in the frontier production function through time. The production function cut with a vertical plane through the origin along the average factor ray. The graphs are based on the estimation for the whole time period 1955-1984.

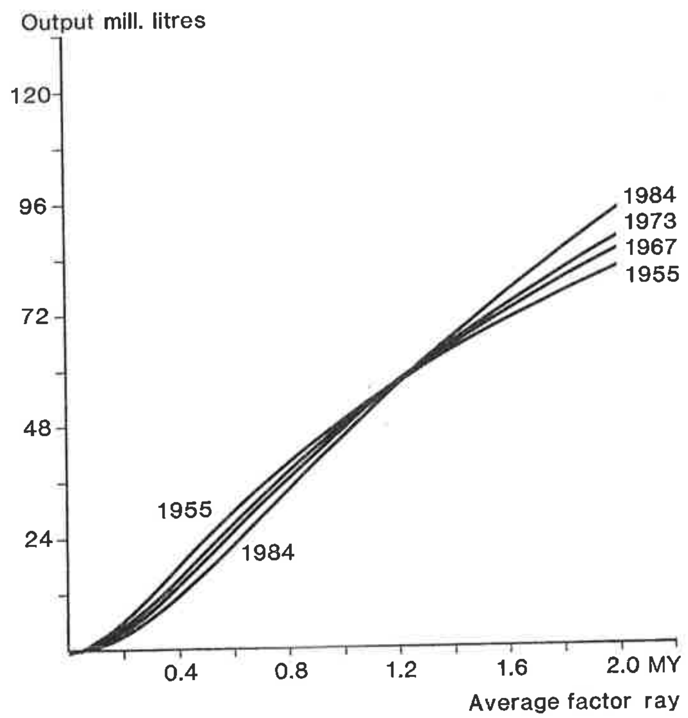


Figure 3.7: The change in the frontier production function through time. The production function cut with a vertical plane through the origin along the average factor ray. The graphs of 1955-1967 and 1973-1984 are based on the estimation by subperiods.

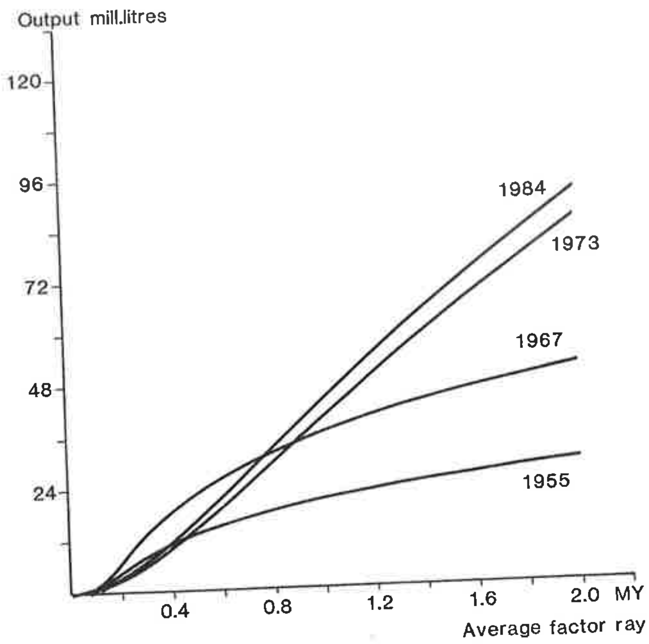
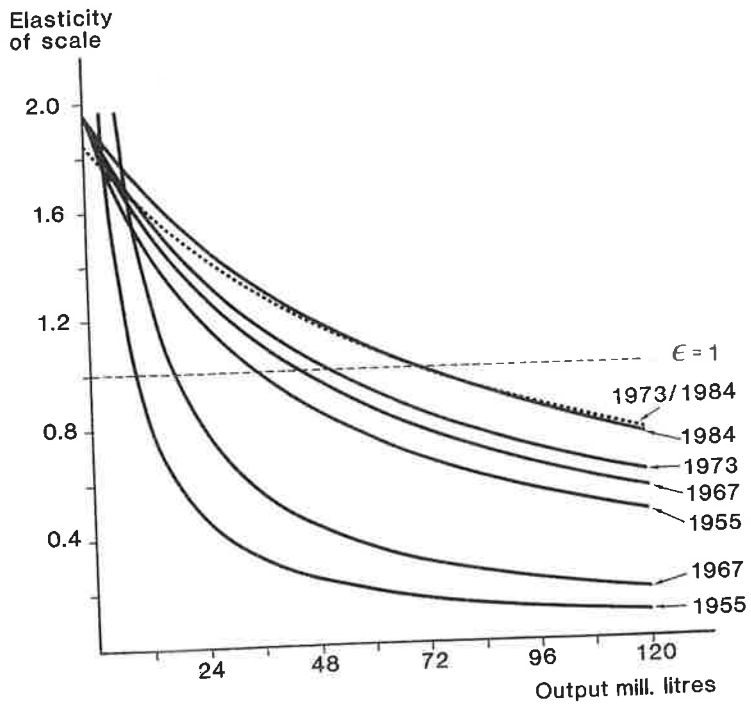


Figure 3.8: Scale elasticity graphs based on the estimation of the whole time period and the subperiods 1955-1967 and 1973-1984.



information about the latest feasible technology for new plants in this output range. Thus in some output ranges the frontier production function may be biased downwards in comparison to a frontier or ex ante production function based on engineering information.

Let us now look at the results of our estimation under the hypothesis of Hicks-neutrality, our Case 2 in Table 3.4. In general the Hicks-neutral results are close to the last year results of the two sub-periods respectively. This holds true both for the kernel elasticities and for optimal scale. Labour has a very high and energy a very low elasticity for 1955-1967, while the opposite is the case for the period 1973-1984. A unique feature of the Hicks-neutral case in 1955-1967 is that there is no trend in the constant term, implying no technical progress during this period. However, the constant term is about three times higher than that in Case 1, reflecting the fact that when the hypothesised specification cannot capture biases and progress, the procedure tries to compensate for this by measuring the "base efficiency", i.e. it increases the function by increasing the constant term and changes the other elasticities. This a feature is also observable in econometric analysis: a wrongly specified estimation brings forth biased estimators. Here we may consider the specification in Case 1 more reasonable than that in Case 2. In fact, the LP problem provides us with a measure of fit namely the value of the objective function i.e. the sum of slacks in our case. It turns out that this value is about 8.6 % higher than in the Hicks-neutral for case 1955-1967, i.e. the fit is much better in the main case than in the Hicks-neutral case. This result coincides with that of Summa (1985) for 1955-1972. On the other hand, the results for the latter subperiods show technical progress under the hypothesis of Hicks-neutrality. In this case the value of the objective function deteriorates only 4.5 % from the main case to the Hicks-neutral one. This is a good example of the importance of dividing the whole period into subperiods corresponding to the main changes in the environment.

3.3.1.3. Further sensitivity of results

The foregoing analysis of the main and Hicks-neutral cases is based on the complete set of time-series cross-section data for 30 years and all plants included in this study in 1955-1984 or in subperiods of 1955-1967 and 1973-1984. We shall now analyse the sensitivity of the results, by excluding some plants from the sample. This is a simple way of testing the sensitivity of the results with regard to the observations. Another, complementary method is to study the behaviour of the cross-section estimates on a year-to-year basis.

Efficient units

The question regarding the impact of extreme observations is rather delicate in the frontier function approach and is often referred to as the "outlier" problem, since the frontier is supported by a subset of the sample and it is difficult to judge in advance whether any of the outliers belong to this subset. Outliers may owe their existence to exceptional technology, scale of production, management performance or faulty data. The choice of method for evaluating the reliability of the frontier estimates must be made on the basis of the purpose of the study and the information gathered about the above factors.

Timmer (1971) tested the sensitivity of the results by discarding efficient units on the frontier from the first run and then reestimating a new frontier without them. According to Timmer's results, the frontier without outliers differed considerably from the "original" frontier. The reestimated frontier came closer to the normal "average" production function, only the constant term differed from the average to any degree. If, for instance, poor quality data produces outliers, it may be an argument in support of Timmer's technique or for the use of a stochastic approach, the latter being more satisfactory from an econometric point of view. But if the extreme observations are caused by efficient production technology and management's X-efficiency, however, discarding the efficient units goes against the principles of the frontier approach.

We performed several sensitivity tests by discarding observations and found it relevant to report on two cases, Case 3, where one of the frontier units is excluded from the data set from 1955-1967, and Case 4, where three frontier plants are excluded from data from 1973-1984. Results from these two sensitivity tests were then compared to the main case. The results are shown in Table 3.4.

In Case 3 the kernel elasticities differ clearly from those of the main case. The capital elasticity gets the value zero. The technical change has been labour saving and energy using. The labour elasticities are roughly of the same magnitude, but the trends have different signs. We also observe a slightly higher optimal scale than in the main case; this would be an expected result if the deleted plant was small.

All elasticities are sensitive to the exclusion of the three frontier plants in our Case 4. The labour elasticity has about the same value in

1973 as in the main case , but due to the opposite signs in trend, the value of Case 4 in 1984 is more than twice the value of the main case. Capital elasticities are higher while energy elasticities are lower. The optimal scale gets the value of 45,100 m^3 at the beginning but increases to 86,400 m^3 in 1984; the former is clearly lower and the latter clearly higher than the optimal scale in the main case. The constant term gets a higher value, but now there is a trend showing technical progress.

Zero parameter estimates are a general feature of the linear programming approach. Had we used a nonlinear method, small nonzero values would probably have been obtained. In the main case for 1973-1984 and in Case 3, the trend in A is zero, which means that the bias completely exhausts the change in the production function, leaving no room for a positive trend in A.

It may also seem puzzling that the trend in α is, in all cases but one in Table 3.4, zero although that of β is positive in two of the four cases when it is allowed to vary. Due to the linear structure of the problem, α and β enter in a symmetrical way it is therefore natural for the entire change in scale elasticity to be captured by changes in one of them.

Cross-section results

For the purpose of analysing the sensitivity of the frontier estimates it is interesting to look at the cross section results reported in Table 3.5 for the LP model, in which we, of course, do not have any trends.

Table 3.5: LP-model estimates for cross-sections 1955-84

$Y^{\alpha} e^{\beta y} = A L^{a_1} KT^{a_2} E2^{a_3}$

Year	Kernel function parameters				Scale function parameters		Optimal scale in m3
	Constant term ln A	Labour elasticity a_1	Capital elasticity a_2	Energy elasticity a_3	α	$\beta \cdot 10^5$	
1955	-1.52	0.51	0.04	0.45	0.43	0.59	9700
56	-4.47	0.50	0.50	0.00	0.20	0.83	9600
57	-2.39	0.68	0.32	0.00	0.38	0.57	10800
58	-2.70	0.48	0.30	0.22	0.37	0.63	10100
59	-3.04	0.60	0.19	0.21	0.27	0.58	12600
1960	-5.06	0.67	0.33	0.00	0.04	0.73	13000
61	-5.11	0.48	0.52	0.00	0.13	0.66	13200
62	0.15	1.00	0.00	0.00	0.51	0.44	11000
63	0.34	1.00	0.00	0.00	0.53	0.41	11500
64	-2.20	0.36	0.00	0.64	0.35	0.45	14500
1965	-5.42	0.24	0.76	0.00	0.00	0.59	17000
66	0.15	0.98	0.00	0.02	0.53	0.28	16700
67	0.09	1.00	0.00	0.00	0.53	0.23	20200
68	-3.42	0.00	0.01	0.99	0.35	0.19	33700
69	-0.43	0.46	0.19	0.35	0.68	0.03	115700
1970	-5.24	0.30	0.51	0.19	0.29	0.14	52100
71	-7.51	0.00	0.64	0.36	0.13	0.18	49000
72	-4.22	0.49	0.00	0.51	0.18	0.17	46900
73	-1.62	1.00	0.00	0.00	0.36	0.13	47600
74	-1.81	1.00	0.00	0.00	0.30	0.20	35400
1975	0.97	1.00	0.00	0.00	0.64	0.00	(1.57)*
76	0.75	1.00	0.00	0.00	0.61	0.00	(1.65)*
77	1.23	0.87	0.13	0.00	0.74	0.00	(1.35)*
78	-5.98	0.69	0.31	0.00	0.07	0.17	55300
79	-6.74	0.65	0.35	0.00	0.00	0.19	51500
1980	-4.19	0.79	0.21	0.00	0.21	0.09	89800
81	-3.06	0.62	0.00	0.38	0.28	0.09	77800
82	-3.78	0.62	0.21	0.17	0.27	0.09	76800
83	-5.01	0.33	0.50	0.17	0.27	0.11	66400
84	-1.39	0.00	0.06	0.94	0.47	0.09	58900

*) Constant elasticity of scale value

In Table 3.5, the kernel function parameters vary a lot but show some conformity with the combined time-series-cross-section estimations: the energy elasticity is rather low, the labour elasticity is high and fairly constant and the capital elasticity is low. The energy elasticities do not react to the first rising step in the prices but do get a clearly positive value in the years 1981-1984, and an exceptionally high value of 0.94 in 1984. The capital elasticity gets a high value during the investment peak in 1970-1971, but returns to zero-level the year after. This could be interpreted as some kind of "overinvesting". The tendency to get zero-elasticities increases, of course, when the number of plants diminishes, as is the case at the end of the observation period.

The scale function parameters behave quite reasonably. The average value over the years 1955-1984 is 0.34, between the values we got in the case of the combined time-series cross-section estimations of 1955-1967 and 1973-1984. The results of the β-parameter vary more but show a strong decreasing trend. The average of the cross-section estimates is close to the value of the main model for both subperiods. As it turns out, the estimates of the cross-sections are, in general, of the "right" magnitude but unstable.

The optimal scale begins at a fairly low level, about 10,000 m^3 in 1955. This is very close to the optimal scale of the first subperiod main model in the same year. The average value of the last five years also matches well with the value of the last subperiod main case.

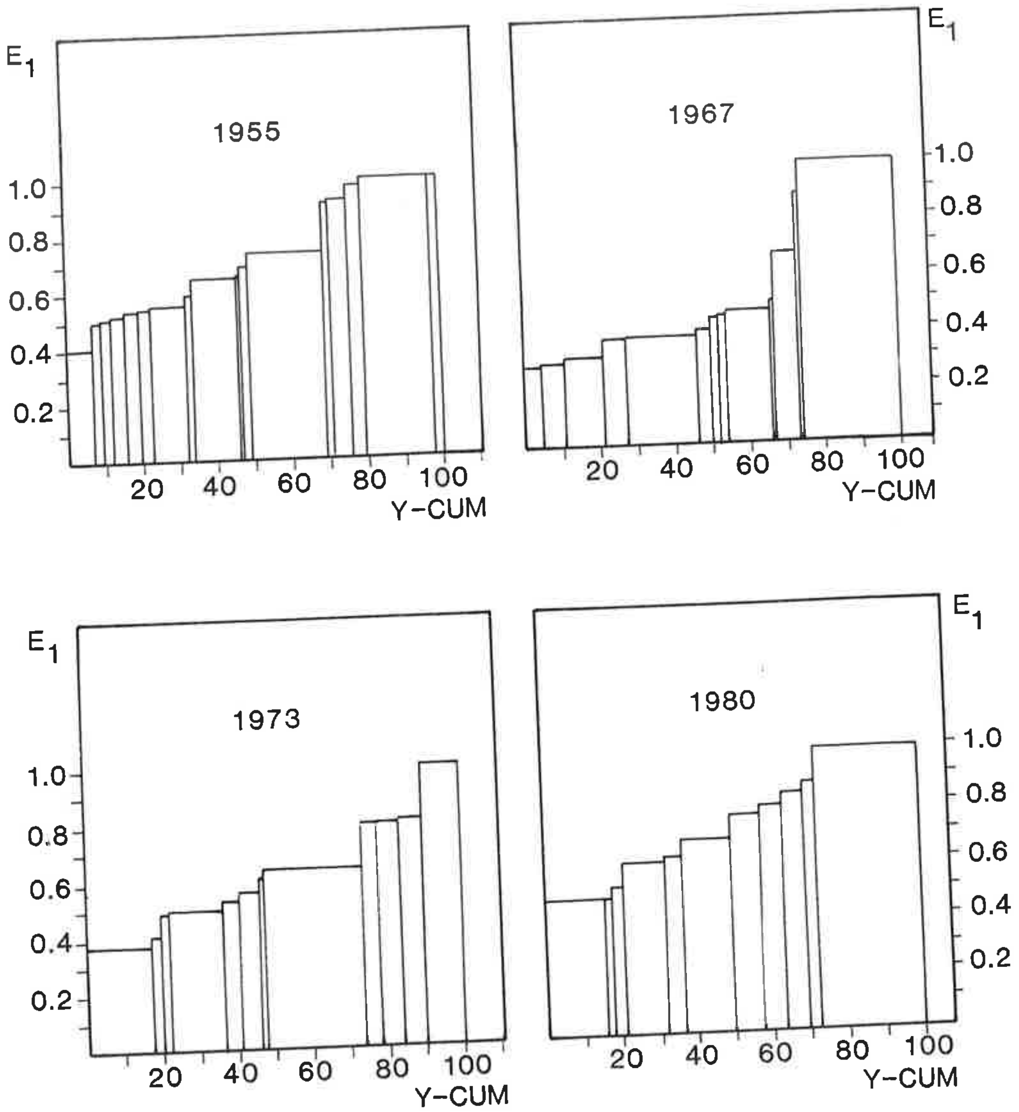
3.3.1.4. Technical efficiency and scale efficiency

We continue the analysis of structural change and technical development using the E measures of plant-specific efficiency introduced earlier in Section 2.3.1. Here the emphasis lies on the efficiency of the micro units. Later, in Chapter 4, we shall look at efficiency at the industry level as a whole.

In Figure 3.9 the units are ranked from left to right in increasing order of their efficiency ratings. Each rectangle represents an individual plant. The figures, analogous to the Salter diagrams, illuminate the range and shape of the efficiency distributions. It is possible to analyse the location of the units of different sizes. We present the results for the years 1955, 1967, 1973 and 1980 based on the main cases of the subperiods in Table 3.4 as we did for the partial productivities in Section 3.2.

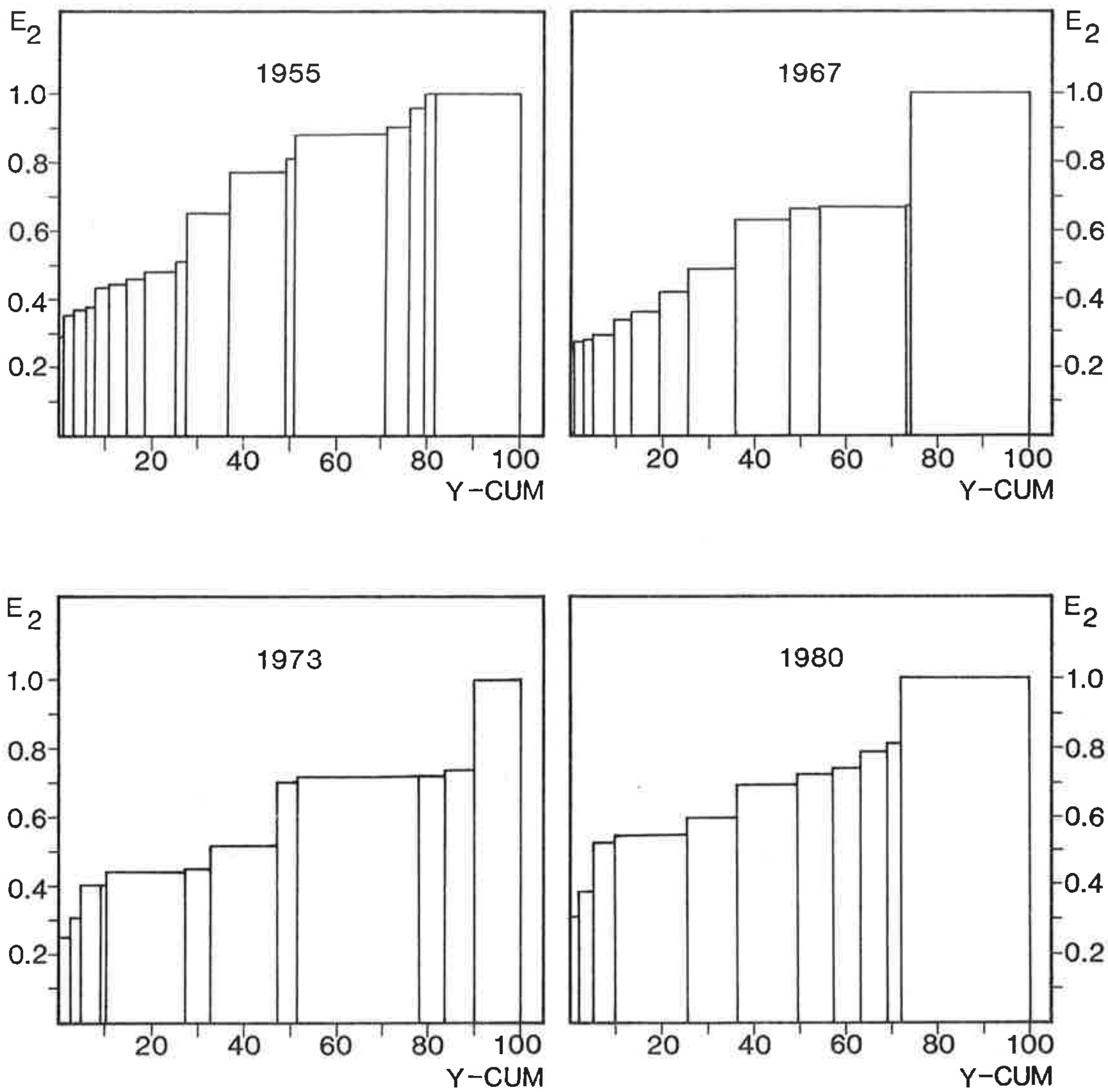
The input saving measure, E_1, shows the ratio between the amount of inputs required to produce the observed output with frontier function technology and the observed amount of inputs. According to E_1 there is a large variation in efficiency between the units in all the years shown. However, in pace with a sharpening competition, the differences have, to some extent, been tapering off. What strikes the eye is how inefficient some relatively large units were in the 1967, 1973 and 1980 and, correspondingly, how evenly the small plants are distributed. At the beginning of the period only five rather small plants seem to be concentrated at the least efficient end of the distribution. Some of the small plants succeeded in benefiting from the strictly regulated market, while the same circumstances limited the efficiency of the largest plants. Under free market conditions in the 1970s and 1980s, there is a distinctly more even distribution between firms of different sizes. In 1955 the least efficient plant had an efficiency value E_1 of 0.4 and produced about 7 % of the industry's total output. The same output could have been produced with only 40 % of the observed inputs by utilising the frontier technology. Until the investment period of the late 1960s, the efficiency value of the least efficient unit gradually decreased and got the value of 0.29 in 1967. In the 1970s the input

Figure 3.9: Development of input saving technical efficiency measure E_1 for selected years.



saving technical efficiency measure E_1 increased gradually and, as a result, the distribution differs clearly from that in 1967. There is also an increase in input saving efficiency for most plants between 1973 and 1980 and in particular for the largest plant.

Figure 3.10: Development of output increasing technical efficienty measure E_2 for selected years.

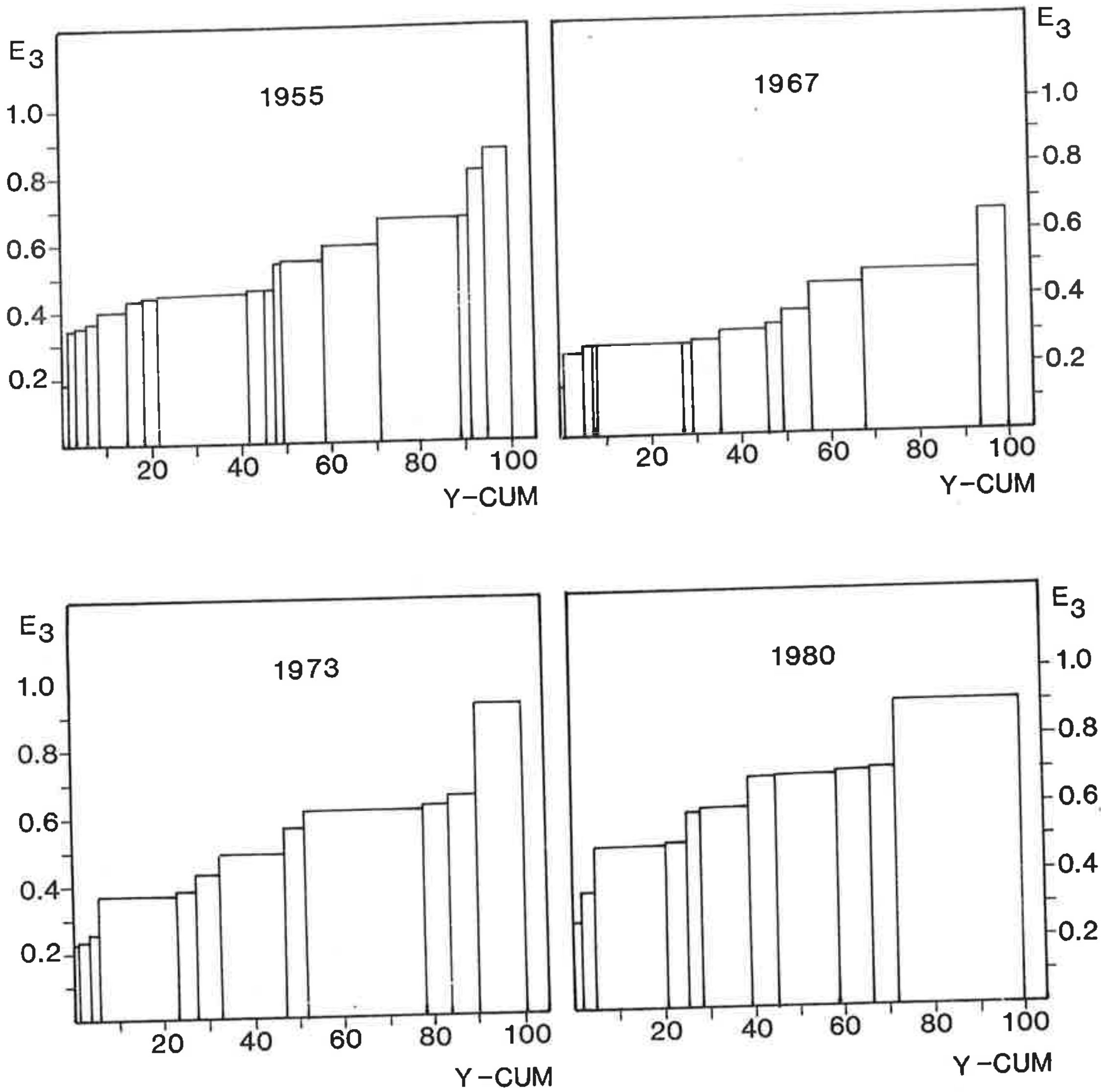


The input saving measure E_1 is relevant if we assume the output to be constant. But if, on the other hand, the amount of key inputs is assumed to be fairly constant the output increasing measure E_2 is the relevant one. As stated earlier, these two measure of technical efficiency will, as a rule, not coincide, except in the case of linear homogeneity. The least efficient unit according to E_2 has an efficiency value of 0.29 in 1955, i.e. the observed production is only 29 % of the output that would be obtained by employing the same amount of inputs in the frontier function. Here, the least efficient plant is small in each year. Moreover, the small plants have a tendency to concentrate at the least efficient end of the distribution.

The rapid increase in optimal scale during the first period leads to a lower level of E_2 values in 1967 except for the most efficient unit which is on the frontier and due to the constant optimal scale level between 1973 and 1984 we get the reverse effect.

In Chapter 2.3.1. we derived three different measures for scale efficiency showing how close an observed plant is to the optimal scale. Here we concentrate on E_3 only, because the other two, E_4 and E_5, are directly derived ratios of E_3, E_1 and E_2 respectively. The least efficient units show a very low scale efficiency indeed. E_3 for the least efficient unit varies between 0.15 and 0.20 before the investment boom and gets slightly higher values after that. Had the least efficient units employed the best-practice technology at the optimal scale level, the level of their potential input coefficients would have been only 15 to 30 % of the actual ones observed. The least efficient unit was one of the three smallest ones in each year. It is worth noting that the scale efficiency values for the most efficient plants also are rather low except for one or two of the most efficient ones. The highest values are, on average, higher after the investment period than before. Small units tend to be the least efficient ones, yet the relationship between size and scale efficiency is not quite as clear as might be expected.

Figure 3.11: Development of scale efficiency measure E_3 for selected years.



We used Spearman's rank correlation coefficient to look at the changes in the ranking order of the individual plants for consecutive panels. In Table 3.6 we present the coefficients between the years and, in addition to these, between the first and last years of the subperiods (1955-1967 and 1973-1984) and of the whole period 1955-1984. For all E, the ranking remains fairly stable in consecutive years until 1982. The smooth, gradual change during the subperiods can be explained in most cases by the minor investment and/or other measures taken by plant management.

Table 3.6: Spearman's rank correlation coefficient between different efficiency measures for selected years.

	E_1	E_2	E_3	E_4	E_5
1955/56	0.94	0.97	0.96	0.97	0.98
56/57	0.94	0.91	0.98	0.97	0.96
57/58	0.84	0.96	0.94	0.96	0.99
58/59	0.93	0.94	0.93	0.99	0.97
59/60	0.89	0.91	0.96	0.98	0.97
60/61	0.97	0.92	0.98	0.99	0.96
61/62	0.87	0.95	0.90	0.96	0.93
62/63	0.96	0.92	0.92	0.97	0.97
63/64	0.89	0.87	0.93	0.99	0.97
64/65	0.93	0.79	0.92	0.98	0.94
65/66	0.95	0.77	0.87	0.96	0.96
66/67	0.82	0.94	0.89	0.96	0.96
73/74	0.83	0.78	0.94	0.97	0.67
74/75	0.77	0.85	0.84	0.97	0.87
75/76	0.94	0.88	0.84	0.99	0.90
76/77	0.62	0.64	0.75	0.98	0.89
77/78	0.82	0.90	0.91	0.99	0.99
78/79	0.86	0.93	0.95	0.98	0.87
79/80	0.66	0.83	0.88	0.92	0.94
80/81	0.86	0.92	0.95	0.96	0.86
81/82	0.85	0.85	0.95	0.88	0.91
82/83	0.55	0.47	0.59	0.69	0.99
83/84	0.89	0.56	0.67	0.98	0.96
1955/67	0.49	0.86	0.63	0.86	0.82
1967/73	0.33	0.28	0.19	0.15	-0.20
1973/84	0.08	-0.03	-0.03	0.62	-0.23
1955/84	0.13	0.38	-0.24	-0.22	-0.28

The rank correlation coefficient between 1955 and 1967 is surprisingly high and in particular for E_2. This is a further indication of a gradually disembodied technical progress influencing all plants more or less uniformly. The period of rapid capacity expansion between 1967 and 1973 yields considerably lower rank correlation coefficients. For E_5 we even get a negative correlation. This may be explained by the great changes in the size structure during this period.

The last subperiod was obviously one of non-uniform technical progress leading to a radical change in efficiency ranking. The results for the whole period 1955/84 also show that it is mostly changes in size structure which have induced the changes in the efficiency rankings.

It is also interesting to look at the consistency of the picture that we get when we use the different efficiency measures. In Appendix 2 we computed the Spearman's rank correlation coefficient between different efficiency measures for selected cross section years.[1)] Most of the correlations are rather high, which means that the same units have a tendency to be efficient/inefficient, no matter which efficiency measure is used. Due to the functional relationships between the efficiency measures, this is, of course, not suprising.

In the opinion of the plant managers interviewed, the efficiency rankings observed and the changes in them were easy to explain. Besides investments, several other factors were mentioned as causes for the changes. These may be summarised under the concept of X-efficiency.

---

1) In Summa (1985) also the Spearman's rank correlations of the labour and capital productivities as well as these of the profitability are included in the comparisons, see pp. 100-101.

3.3.1.5. Generalised Salter measures

In Section 2.3.3. we introduced different ways of measuring the impact of technical change in the frontier function framework. Here we extend the decomposition of the Salter measure of technical advance to three inputs in the homothetic production function case. Then we calculate Salter's and Binswanger's bias measures of technical advance.

Førsund and Hjalmarsson (1979) have shown that the rate of technical advance measure, T, between two time points t_0 and t_1 can be split into two main components: the reduction in unit costs due to the movement along a factor ray (T_1) and the reduction in unit costs due to the movement along the next period's efficiency frontier (T_2). In the case of a homothetic production function, the unit cost reduction due to movement along a factor ray T_1 can be further decomposed into a reduction in unit costs due to a change in the optimal scale (OS), cost reduction due to Hicks-neutral technical progress (H), and cost reduction due to a factor-bias technical change for a constant factor ratio (B) and thus T_1=OSxHxB.

With the homothetic production function specified above and using the formulas derived in the Appendix of Førsund and Hjalmarsson (1979), extended to three inputs and simplified, the following expressions for the change in unit costs at the optimal scale from year t_0 to t_1 are obtained (where, $t_1 - t_0 = \theta$):

$$OS = \frac{\left[\frac{e\ \beta(t_1)}{1 - \alpha(t_1)}\right]^{[1-\alpha(t_1)]}}{\left[\frac{e\ \beta(t_0)}{1 - \alpha(t_0)}\right]^{1 - \alpha(t_0)}}$$

$$= \frac{\left(\frac{e(\beta-\gamma_6(t+\theta))}{1 - (\alpha-\gamma_5(t+\theta)}\right)^{1-(\alpha-\gamma_5(t+\theta))}}{\left(\frac{e(\beta-\gamma_6 t)}{1 - (\alpha-\gamma_5 t)}\right)^{1-(\alpha-\gamma_5 t)}}$$

$$H = \frac{A(t_0)}{A(t_1)} = e^{-\gamma_4 \theta}$$

$$B = \prod_i x_i^{a_i(t_0) - a_i(t_1)} = x_1^{\gamma_1 \theta} \cdot x_2^{\gamma_2 \theta} \cdot x_3^{\gamma_3 \theta}$$

$$T_2 = \prod_i \left[\frac{a_i(t_0)}{a_i(t_1)}\right]^{a_i(t_1)}$$

$$= \prod_{i=1}^{3} \left(\frac{a_i - \gamma_i t}{a_i - {}_i(t+ \ )}\right)^{[a_i-\gamma_i(t+\theta)]}$$

The empirical results are presented in Tables 3.7 and 3.8. First we start with a discussion of the technical advance results for the estimated periods in Table 3.7 and then we discuss the development between 1967 and 1973 in Table 3.8.

For the first subperiod the overall technical advance measure T is 0.56, i.e. the average cost at optimal scale in 1967 is 56 % of the average cost at optimal scale in 1955, representing a strong annual decrease in the average cost of 4.89 %. During the second subperiod the rate of technical progress is very slow and the annual decrease in average cost at optimal scale is below one per cent. These results are consistent with the visual impression from Figure 3.7. The result obtained for the whole period also indicates a fairly slow rate of technical advance, again due to the fact that the results for the whole period are similar to those of the last subperiod.

The decomposition of the overall measure shows that the factor bias advance T_2 is of less importance during both subperiods. On the other hand, the change in optimal scale contributes a lot during the first subperiod. The cost-reducing effect of the constant term is exactly balanced by the cost-increasing effect of the proportional change due to bias.

During the second subperiod, optimal scale is constant and neither OS nor H contributes to technical progress. it is the bias terms that lowers the average cost.

The result for the whole period deviates considerably from those of the subperiods. In this case the cost saving impact of factor bias advance, T_2 is dominating while T_1 in fact yields a negative contribution to growth. This result further confirms our view that a single function is not able to yield a reliable picture of the growth process for such a long period containing different phases of structural change.

One might suspect that these measures are sensitive to the chosen factor ratio, but sensitivity analyses show that variations in the factor ratio have a fairly small influence on the degree of unit cost changes for these periods. We will return to this point below.

Table 3.7: The Salter measure of technical advance and its components along the average 1955-84 factor ratio. Annual percentage changes in parentheses.

Type of relative unit cost reduction		1955/67	1973/84	1955/84
T:	Overall technical advance	0.56 (4.89)	0.91 (0.73)	0.61 (1.71)
T_1:	Proportional technical advance	0.61 (4.15)	0.92 (0.73)	1.05 (-0.18)
OS:	Change in optimal scale	0.61 (4.98)	1 (0)	0.72 (1.15)
B:	Proportional change due to bias	2.42 (-7.36)	0.92 (0.73)	1.47 (-1.33)
H:	Hicks-neutral advance	0.41 (7.36)	1 (0)	1 (0)
T_2	Factor bias advance	0.91 (0.75)	0.98 (0.19)	0.58 (1.90)

Table 3.8: The Salter measure of technical advance and its components along average factor ratios. Annual percentage changes in parentheses.

Type of relative unit cost reduction		1955/67	1973/84	1967/73
T:	Overall technical advance	0.42 (14.55)	0.34 (17.88)	0.38 (16.11)
T_1:	Proportional technical advance	1.42 (-5.86)	1.16 (-2.53)	1.29 (-4.30)
OS:	Change in optimal scale	10.66 (-39.44)	10.66 (-39.44)	10.66 (-39.44)
B:	Proportional change due to bias	0.80 (3.77)	0.65 (7.11)	0.73 (5.34)
H:	Hicks-neutral advance	0.17 (29.81)	0.17 (29.81)	0.17 (29.81)
T_2	Factor bias advance	0.29 (20.41)	0.29 (20.41)	0.29 (20.41)

It is particularly interesting to investigate the change from 1967 to 1973 by comparing the end year parameter results for the first subperiod with the first year parameters of the second one. This is done in Table 3.8 for technical advance and in Table 3.9 for bias. Since the B-component is a function of the chosen factor ratio we have calculated B for the average factor ratio for the entire period as well as for the two subperiods.

In this case B is fairly sensitive to the chosen factor ratio but we should observe that the factor ratio in 1973/84 differs a lot from that in 1955/67 since the investment boom had a strong impact on the average factor ratio and the changes in kernel elasticities from 1967 to 1973 are also fairly strong. But even if the size of the B component differs between the factor ratios it is important to note that even at the 1955/67 average factor ratio B contributes markedly to the reduction in unit costs which indicates a shift in technology and not just substitution.

As regards the overall level of technical advance there has been an extraordinarily strong technical progress between 1967 and 1973 in the interval of about 14-18 % annual reduction in unit costs. It is particularly the Hicks-neutral and factor bias advance components which contribute to the strong progress. As one might suspect from a look at Figure 3.7 OS contributes to a strong negative progress which are offset by the other components. The result for this period should be taken as a strong indication of the importance of embodied technical progress though it is not easy to separate the embodiment effect from the effect on productivity growth of the simultaneous deregulation of the beer market.

Let us now look at the of technical advance bias. The general version of the Salter bias measure, introduced earlier, is in our case given by[1)]

$$D_{ik} = \frac{a_{i,t_1}}{a_{i,t_0}} \cdot \frac{a_{k,t_0}}{a_{k,t_1}}$$

and the Binswanger cost share measure is simply given by[2)]

$$S_i = \frac{a_{i,t_1}}{a_{i,t_0}}$$

The Salter measure is a relative measure of bias in relation to a certain factor and shows the relative change in the optimal factor ratio for the two factors under consideration. It is easy to see that $D_{ik} = S_i/S_k$. The results are presented in Table 3.9.

The absolute Binswanger bias measures imply a capital and energy saving but labour using technical progress during the first subperiod in the frontier production function. During the second subperiod this bias changes to labour and capital saving but energy using which is also the result obtained for the whole period as well as for the jump from 1967 to 1973.

As regards the relative Salter measures it turns out that bias is labour using relative to capital and capital saving relative to energy in all cases.

---

1) For the derivation see Førsund and Hjalmarsson (1979).

2) For the derivation, see Binswanger (1974b) and Førsund and Hjalmarsson (1984).

Table 3.9: Bias measures of technical advance.

Type of bias measure	1955/67	1967/73	1973/84	1955/84
Salter				
$D_{L,KT}$	2.67	2.18	1.33	1.22
$D_{L,E2}$	1.59	0.04	0.61	0.11
$D_{KT,E2}$	.59	0.02	0.46	0.09
Binswanger				
S_L	1.39	.32	0.69	0.62
S_{KT}	0.52	.15	0.52	0.51
S_{E2}	0.87	7.99	1.12	5.75

This means that for constant factor prices it would be optimal to increase the labour-capital ratio (and decrease the capital-energy ratio). This does not in reality imply, however, that the labour-capital ratio will be increased since the actual choice of technology also depends upon the expected development of factor prices.

3.3.2. Stochastic approaches

Since the relevance of applying different approaches to frontier function estimation is intensively discussed in the literature[1)], we here

---

1) For a recent survey see Schmidt (1985).

also complement the deterministic approach with the stochastic one. In Chapter 2 we discussed the basic idea of the stochastic approach and here the same kind of models will be applied.

Firstly, the deterministic results of our main case are compared with the corresponding results of the composed error model with normal - half normal error distribution. Secondly, to further illustrate the differences between the deterministic and the stochastic approach, we also look at the Cobb-Douglas production function, since this is the model used most frequently in the literature.[1] The main results are set out in Tables 3.10 and 3.12.

The model is:

$$y = f(x) \exp(v - u)$$

presented earlier in Chapter 2, where v is symmetrically distributed and able to capture the random effects or statistical noise which cause variation in the placement of the deterministic kernel f(x) across firms. The other component u has a one-sided distribution and captures the technical inefficiency relative to the stochastic production frontier, due to failure to produce the maximum possible output using a given set of inputs. This model, introduced by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) can be seen as an extension of the research with emphasis on the skewness of the error distribution.[2]

---

1) See e.g. Greene (1980).

2) See e.g. Aigner, Amemiya and Poirier (1976).

In all stochastic frontier models to date, the errors representing statistical noise are assumed to be iid normal. A number of distributions have been assumed for the one-sided (inefficiency) errors. Firstly we use the most common assumption, that of half-normal distribution.[1)]

If the two errors are assumed independent of each other and of the inputs, and specific distributional assumptions are made (e.g. normal and half-normal, respectively), then the likelihood function can be defined and MLEs can be calculated. This will - generally - require a numerical maximisation of the likelihood function.[2)]

The stochastic composed error model with a normal - half-normal error structure did degenerate, in the sense that it was not possible to distinguish the stochastic frontier from the average function based on OLS estimation. Thus the OLS estimates also represent the stochastic frontier in this case. The reason behind this result is that if the efficiency distribution is symmetric, then no difference between the stochastic and the normal OLS is found.

The deterministic LP results and the OLS results are presented in Table 3.10. In order to make the OLS-results as comparable as possible with the LP-frontier we have imposed restrictions on some parameters, i.e. which had the wrong sign in the first run without any restrictions. It

---

1) Stevenson (1980) generalised half-normal by considering normal (not necessarily with zero mean), truncated from below at zero, while Lee (1983) has considered four-parameter Pearson family of distributions. The latter has considerable generality, at the price of considerable complexity.

2) Computational issues are discussed by Waldman (1982), Greene (1982), Lee (1983) and Huang (1984).

Table 3.10: Estimates of the deterministic frontier production function and the average (OLS) model (Main case). Standard errors in parentheses.

$$Y^{\alpha-\gamma_5 t}\, e^{(\beta-\gamma_6 t)y} = Ae^{\gamma_4 t}\, L^{(a_1-\gamma_1 t)}\, KT^{(a_2-\ _2 t)}\, E2^{(a_3-\gamma_3 t)}$$

	Deterministic (LP)			Average (OLS)		
	1955-67	1973-84	1955-84	1955-67	1973-84	1955-84
ln A	-4.19	-1.27	-1.98	-1.82 (1.02)	1.51 (0.61)	.33 (.30)
γ_4	0.07	0	0	0.27 (0.12)	0*	0*
a_1	0.49	0.24	0.46	0.69 (0.16)	.84 (0.36)	.93 (1.24)
γ_1	-0.02	0.01	0.01	-0.03 (0.00)	.01 (0.01)	0.01 (0.01)
a_2	0.40	0.03	0.47	0.12 (0.11)	.16 (0.07)	0*
γ_2	0.02	0.00	0.01	0.02 (0.01)	-0.001 (0.01)	0*
a_3	0.11	0.79	0.07	0.19 (0.16)	0*	.07 (.09)
$\gamma_3 \cdot 10$	0.00	-0.01	-0.14	0.01 (0.02)	0*	-.01 (.01)
α	0.20	0.54	0.51	0.45 (0.10)	.49 (0.06)	.64 (.03)
γ_5	0	0	0	-0.02 (0.01)	-.004 (0.005)	0.00 (0.00)
$\beta \cdot 10^5$	8.37	0.64	1.37	4.35 (1.16)	0.11 (0.02)	22.4 (3.2)
$\gamma_6 \cdot 10^7$	31.05	0	2.29	22.52 (11.70)	0.52 (0.23)	0.08 (0.01)
Optimal scale in m^3	9 900- 18 500	72 000	36 000- 71 700	12 900- 21 800	50 400- 96 200	29 600- Increasing towards infinity

Figures marked by * are restricted.

was not possible to impose all restrictions used in the deterministic case since the problem did not converge in that case.

Looking first at the standard errors of the OLS estimates the results show that these were in most cases fairly small except for the trends in the marginal elasticities. It also turns out that the point estimates of the marginal elasticities in the OLS results differ considerably from the deterministic frontier results. In general the marginal elasticity of labour is in the average function higher than at the frontier and the trends have about the same sign and magnitude. For capital and energy there is no systematic pattern in the differences at all.

As regards scale properties, optimal scale levels are fairly similar during the first subperiod. During the second subperiod we get an increase in optimal scale for the average function but a constant optimal scale at the deterministic frontier. For the whole period optimal scale increases rapidly towards infinity. However, since infinity is reached before the end of the period we get a negative value for optimal scale during the last years due to a negative sign for the β-parameter.

The shape of the average and frontier production functions for the two subperiods are shown in Figures 3.12 and 3.13.

During the first subperiod between 1955 and 1967 the distance between the average function and the frontier function grows strongly while this distance shrinks strongly between 1973 and 1984. In the last year the average function even crosses the frontier though at a rather high output level. In 1984 the largest plant produces about 100 000 m^3 and is close to the frontier function.

Figure 3.12: The surface of the frontier (F) and average (A) production function 1955 and 1967 cut with a vertical plane through the origin along the average 1955/84 factor ray.

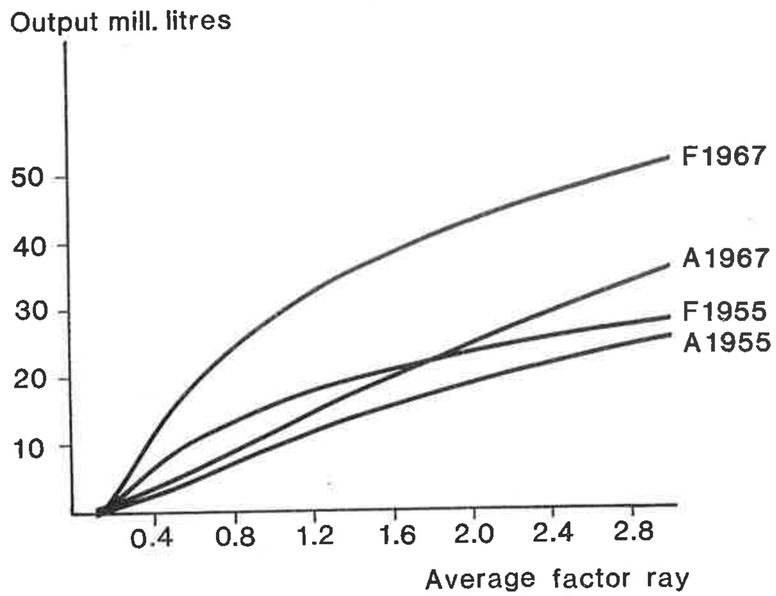
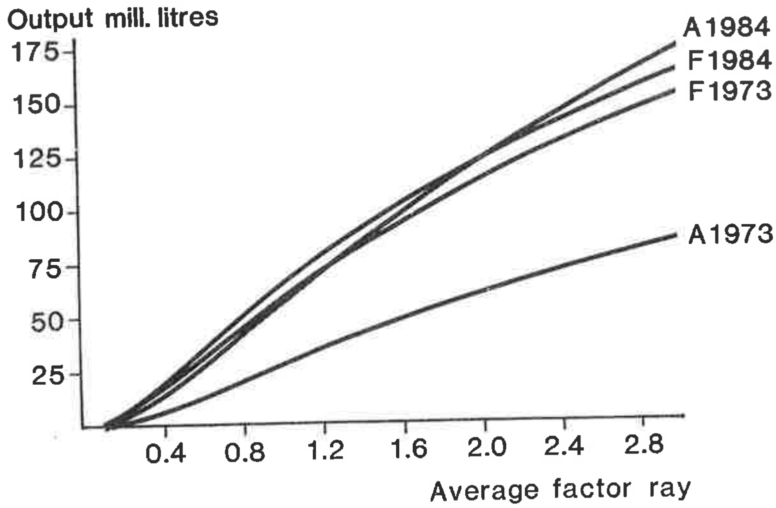


Figure 3.13: The surface of the frontier (F) and average (A) production function 1973 and 1984 cut with a vertical plane through the origin along the average 1955/84 factor ray.



These results are interesting. They indicate that during the period 1955-67 there was new technology available at the frontier but this technology was not adopted by the firms to any large degree. Therefore the average is lagging behind the frontier. Between 1973 and 1984 the best-practice frontier moved slightly upwards while most firms succeeded in improving their efficiency considerably. These results are consistent with the radical change in competition that ocurred between the two periods.

We have also calculated Salter's measure of technical advance for the average function: The comparison is limited to the two subperiods, see Table 3.11.

It turns out that during the first period technical progress is much slower in the average function than on the frontier, but during the second period the opposite holds, consistent with the visual impression in Figures 3.12 and 3.13 above. This should yield a decreasing structural efficiency during the first period and an increasing structural efficiency during the second period which also is confirmed below in Section 4.1.

There are also other distributions apart from the half-normal one which do meet the requirements of ML estimation.[1] A particularly attractive one for the frontier estimator is the gamma density

$$f(u) = G(\lambda, P) = \frac{\lambda P}{T(P)} u^{P-1} e^{-u}, \quad u \geq 0,\ \lambda > 0,\ P > 2$$

---

1) For detailed presentation of the requirements see Greene (1980), pp. 34-41. For the estimation of the gamma function we have used Greene's program in Göteborgs Data Central.

Table 3.11: The Salter measure of technical advance and its components along the 1955/84 average factor ratio. Annual percentage changes parentheses.

Type of relative unit cost reduction		1955/67	1955/84	1973/84
T:	Overall technical advance	- *	0.99 (1.4)	0.51 (6.1)
T_1:	Proportional technical advance	1.08 (-0.6)	0.98 (1.9)	0.55 (5.5)
OS:	Change in optimal scale	5.65 (-14.4)	1.16 (-15.9)	1.15 (-1.2)
B:	Proportional change due to bias	4.79 (-12.1)	1.12 (-12.0)	0.48 (6.7)
H:	Hicks-neutral advance	0.04 (27.1)	0.76 (23.8)	1 (0.00)
T_2	Factor bias advance	- *	1.00 (0.4)	0.93 (0.7)

*) A negative marginal elasticity after 6 years yields a negative value of T_2.

The mean and the variance of u are $\mu = P/\lambda$ and $\delta^2 = P/\lambda^2$. The ancillary parameters λ and P provide additional information on the shape of the distribution with which we may characterise our observations as regards relative efficiency and offer some evidence on the relationship between the frontier and average estimators.

For the general case of $G(\lambda,P)$, P must be positive. For the gamma density with P > 2, maximum likelihood estimation of the parameters is a regular case.[1] As $P \to \infty$, the distribution of u tends to normality. This implies that the ML estimator should approach the OLS estimator.

1) See Greene (1980).

The last-mentioned property makes the gamma density attractive for estimating the production frontier, as it implies that the model is quite flexible with regard to the shapes of the error distributions it will accommodate. The gamma specification allows a relationship to be established between the average, which is generally understood to be OLS, and frontier estimators. If the disturbances about the frontier estimator tend to be symmetrically distributed we should expect the average estimator to be a displaced or simply scaled version of the former but with the same shape. The more skewed the disturbances about the frontier are, the less likehood that the frontier estimator resembles the 'average' estimator.

Stochastic frontier models have typically assumed fairly restrictive functional forms, such as the Cobb-Douglas. So in Tables 3.12 and 3.13 we present the results of the stochastic model with gamma distribution in the Cobb-Douglas case and compare the results with the deterministic and OLS estimates. Even in the Cobb-Douglas case it turned out that the normal - half-normal composed error model degenerated to the average production functions.

Since the gamma distribution is assymmetric ML estimation should be more efficient than OLS and we should expect the gain in efficiency obtained by ML to be related to the degree of skewness of the distribution. The skewness coefficient, $E(u-E(u))^3/\sigma^3$ is obtained as $2/\sqrt{P}$, and a large P implies a symmetric distribution. In such a case we should expect the frontier function to be a scaled version of the average function with the same shape. Thus, the P parameter is crucial and we should expect the greatest efficiency gain when P is small (near 2). According to Greene (1980) the value of P/(P-2) should indicate the relative asymptotic efficiency of ML over OLS.

In our cases the value of P is moderately large and far above 2 so the skewness coefficient is fairly small and therefore the efficiency gain of ML over OLS is moderate, in the range of 1.20 to 1.30. Thus we should expect fairly similar results between the average and the composed error estimates.

In Table 3.12 we make a comparison between Cobb-Douglas production functions with trends in all parameters. We have not tried to impose any restrictions on the parameter values in the stochastic cases. The comparison is limited to the two subperiods for which we should expect reasonable results. Since the purpose here is a comparison between different methods of frontier estimation we are not elaborating the economic implications of the results in detail.

The overall impression one gets from the estimates in Table 3.12 is that the results are not very reasonable from the economic point of view. For all but one of the marginal elasticities and their trends the standard errors are large in the first period but considerably smaller during the second period. As regards the estimated parameter values, the trend in the constant term is large in all cases. Furthermore the marginal elasticities are often of the wrong sign and in the last period while the magnitudes of the labour and energy elasticities are quite unreasonable.

There is, in both periods, considerable similarity between the composed error model results and those of OLS except for the constant term. In most cases the asymptotic standard errors in the composed error model are of the same magnitude but slightly smaller than those of the average model. First period estimates show some similarity between the stochastic results and the LP results.

From an economic point of view the results in Table 3.13 are more reasonable. All technical progress is now represented by the change in the constant term. In most cases the standard errors are fairly small and the parameter values are in most cases reasonable. But even in this case the differences between the stochastic frontier and the average function estimates are very small. However, the LP frontier differs considerably from the stochastic frontier, in particular between 1973 and 1984. First period estimates for technical progress are lower using the deterministic frontier approach than in the stochastic frontier case and in that of the average function. Using the deterministic frontier model to estimate second period technical progress, we observed a markedly higher constant term than for the other functions.

It is also interesting to compare the average efficiency level E(u) obtained for the stochastic frontier with the results from deterministic case in Appendix 3 for structural efficiency. For the deterministic frontier the value of structural efficiency is higher in the last period in comparison with the first while the opposite holds for the stochastic frontier. Moreover, the efficiency values for the stochastic frontier are much higher than for the deterministic frontier particularly for the first period.

The main conclusions from this comparison are:

1. The difference between the average function and the stochastic frontier is small in all cases
2. The average level of efficiency for the deterministic frontier is very high in comparison with the level of structural efficiency

Table 3.12: Estimates of the deterministic (LP) and gamma frontier production functions and the average (OLS) model with trends in all parameters (Cobb-Douglas case). Standard errors in parentheses.

	1955-1967			1973-1984		
	LP	Gamma	OLS	LP	Gamma	OLS
ln A	4.446	3.334 (.545)	2.439 (.592)	-.311	10.661 (2.203)	9.950 (2.403)
Trend A	.471	.216 (.073)	.215 (.083)	.201	-.318 (.089)	-.308 (.098
a_1	.878	1.106 (.157)	1.103 (.179)	1.706	4.361 (.557)	4.358 (.611)
Trend a_1	-.121	-.029 (.019)	-.027 (.021)	.045	-.127 (.022)	-.138 (.024)
a_3	0	.165 (.173)	.158 (.197)	0	-3.600 (.640)	-3.629 (.702)
Trend a_3	0.068	.040 (.020)	.037 (.023)	-.037	-.135 (.026)	-.146 (.028)
a_2	0	-.082 (.110)	-.081 (.126)	0	.143 (.336	.144 (.369)
Trend a_2	0.063	.006 (.014)	.007 (.016)	.017	-.017 (.014)	-.004 (.016)
λ		9.278 (2.198)			14.640 (5.132)	
P		8.548 (3.500)			11.858 (7.499)	
E(u)		0.921 (0.170)			.810 (2.38)	
Var(u)		0.099 (0.011)			0.55 (.007)	
$2/\sqrt{P}$		.684			.581	
P/(P-2)		1.305			1.203	

Table 3.13: Estimates of the deterministic (LP) and gamma frontier production functions and the average (OLS) model (Cobb-Douglas case). Standard errors in parentheses.

	1955-1967			1973-1984		
	LP	Gamma	OLS	LP	Gamma	OLS
ln A	5.540	4.460 (.312)	3.532 (.300)	3.323	1.264 (.394)	.438 (.377)
Trend A	.025	.035 (.006)	.035 (.006)	0	.064 (.006)	.064 (.007)
a_1	.727	1.272 (.069)	1.271 (.079)	.206	.826 (.088)	.826 (.099)
a_3	0	-.100 (.076)	-.102 (.087)	.810	.114 (.097)	.112 (.109)
a_2	0	-.060 (.055)	-.106 (.063)	0	.243 (.055)	.279 (.062)
λ		9.224 (2.189)			11.745 (3.698)	
P		8.572 (3.517)			9.6898 (5.371)	
E(u)		0.929 (0.172)			0.825 (.208)	
Var(u)		0.101 (0.011)			0.070 (.010)	
$2/\sqrt{P}$		.683			.642	
P/(P-2)		1.304			1.26	

These results illustrate rather well the main weakness of the stochastic frontier approach: In the case of a fairly symmetric error distribution the stochastic frontier is very close to the average function and a large share of productive inefficiency is explained by stochastic phenomena.

3.4. Short-run industry production function (SRIPF)

3.4.1. Calculation procedure

In Section 2.4 we have defined the SRIPF, described its properties and usage and shown its potential in deriving several concepts of technical and structural change. In this empirical part we, after explaining the calculation procedure, report on the development of region of substitution, cost functions, substitution and scale properties and, finally, technical advance and bias.

In the SRIPF the ex post micro production functions are aggregated in an efficient way to an industry production function. Thus, the capacity distribution is aggregated in an efficient way, to a production function in the output-input space. For each amounts of inputs the output level is maximised. This corresponds to sweeping a quasi-rent line outwards through the capacity distribution in the input coefficient space and operate all units with nonnegative quasi-rents. Since the SRIPF is a non-parametric production function it is often more enlightening to represent the function by its substitution region and a suitable number of isoquants. In our case we put labour on the abscissa axis and energy on the ordinate axis.

The boundaries of the substitution region are easily obtained by a simple ranking of the units according to their input coefficients for each output at a time. In our case, along the upper boundary, the units are taken into operation according to decreasing labour productivity and along the lower boundary according to decreasing energy productivity. In relative price terms this means zero price of energy along the upper boundary and zero labour price along the lower boundary when sweeping the quasi-rent line through the capacity distribution.

If the production units have about the same ranking in both dimensions for both inputs we get a very narrow substitution region but if the production units are scatterad in a north-west - south-east direction in the capacity distribution the substitution region is wide. Several factors determine the shape of the substitution region e.g. the development of the ex ante function in the past and relative price expectations at the time when the technology choice was made.

In the interior of the substitution region the capacities of the micro units are combined into efficient combinations. Since an individual micro unit may be represented by a vector the SRIPF is made up of all efficient combinations of such vectors. Graphically this means that the substitution region is made up of parallelograms where each parallelogram is an efficient combination of two micro units. An example is shown in Figure A.2. in Appendix 4.

The construction of an isoquant is a fairly complicated procedure to explain simply. We have therefore added a specific example detailing the working of the algorithm in Appendix 4, but here we try to give an intuitive explanation of this construction.

Let us start at the upper boundary of the substitution region and at a certain output level. At this output level a unit is in general partly utilised and the question is which micro unit should be activated along the first isoquant segment together with the start unit at the boundary. Along an isoquant segment the capacity utilisation of one unit increases while that of the other unit decreases. The idea behind the selection of the new unit is to compare the angles between the starting unit and all other units in the input coefficient space and to pick out, among the set of negative angles, the unit yielding the steepest angle of the first isoquant line segment. The first segment cannot be steeper than vertical, because the boundary units were arranged in an increasing order of their abscissa input requirements, equivalent of arranging them in an increasing order of abscissa input coefficient. The actual length of the segments dependends on the capacity and input coefficients of the activated unit. On a given isoquant line segment, two units at most can be partly utilised.

The first corner point is reached either when the capacity utilisation of one of the units reaches zero or 100 %. The next task is to compare the angles of all other units in the input coefficient space with the partially activated unit at the corner point found above. The angle of the line segment is then determined by the unit giving the steepest angle next to the one of the previous line segment, etc. until the lower boundary is reached.

The successive angles, in the input coefficient space, for the connecting lines between the units activated along the isoquant are identical with those of the slopes of the line segments in the input space. The isoquant obtained according to the algorithm described above is piecewise linear, convex and is as "close" to the origin as possible. This procedure may be repeated for other industry output levels and thus the graph of the short-run industry production function may be drawn at a desired density.

3.4.2. Development of the short-run industry production function

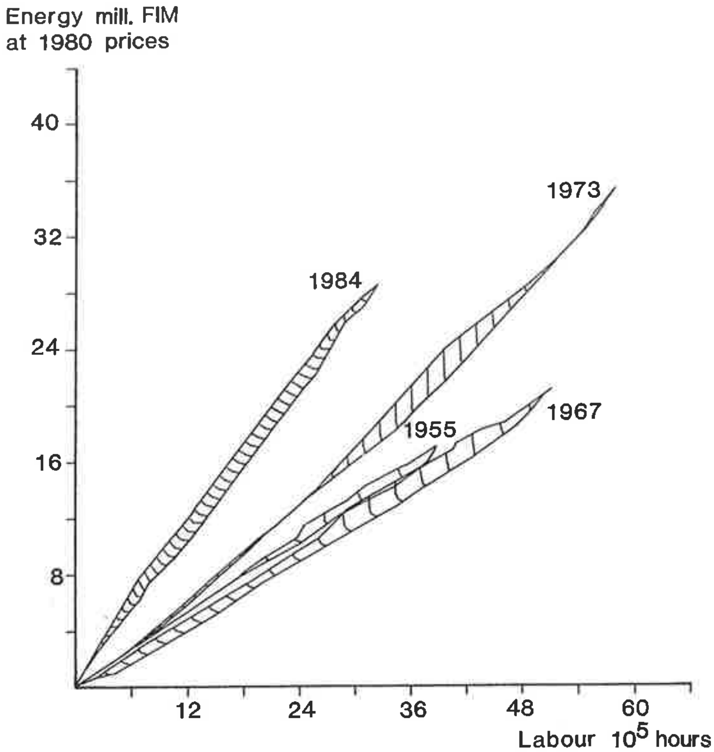
The region of substitution and the isoquant map of the short-run industry production function is presented in Figure 3.14 for the selected years 1955, 1967, 1973 and 1984.

It may also be valuable to study the capacity region which is the transformed isoquant map of the short-run function. This shows the region of feasible input coefficients of the industry production function as a whole. Thus, this region must necessarily be narrower than the area of the capacity distribution of the individual units. The boundary towards the origin of the feasible region is called the efficiency frontier. The development of the capacity region is shown in Figure 3.15.

As stated earlier, there are many factors determining the timing, factor proportions and scale of investments; the expected future development of prices of the major inputs, the ex ante technology and, especially in our case, the changes in demand are usually the most important ones. The properties of the new capacity and the rate at which old machinery is scrapped are the forces forming the shape of the SRIPF.

Comparing the substitution regions for different years, we see from both figures that the width of the region of substitution varies a lot both between different parts of the substitution region within a year and across years. The substitution region is particularly narrow for fairly high output levels both in 1955 and 1973 while there are larger substitution possibilities at most output levels in 1967 and 1984. In Figure 3.14 we see that the static period of 1955-67 only shows a slight movement of the substitution region towards the labour axis; technical

Figure 3.14: The development of the short-run industry production function between 1955 and 1984. The distance between the isoquants is 10 000 m^3.



progress is almost neutral. It is also interesting to note that the average energy-labour ratio was in fact higher in 1955 than in 1967 in spite of the fact that the energy-labour price ratio was 60 % higher in 1967 in comparison with 1955. The same conclusion can be reached when we compare the years 1973 and 1984 since the relative price of energy is slightly higher in 1984 than in 1973. The reasons for these biases will be discussed below in connection with the Salter bias measures.

During the heavy investment period, 1967-1973, there was a dramatic shift towards the more energy intensive direction. This development corresponds well with the impression one gets from the changes in the capacity distributions.

The isoquant levels and the scale of the axis remain unchanged throughout the period, making it possible even to study the movements of the isoquants. The number of the isoquants indicates the total volume produced by the industry in a particular year.

Changes in productivity can be studied by following the movement of isoquants representing the same output levels. Since it is not easy to follow a particular isoquant level in Figure 3.14 we will calculate the advance in the technical progress between the comparison years by the Salter measures below.

By analysing the shape of the isoquants inside the capacity region or the SRIPF, it is possible to study the substitution properties. There is, of course, by assumption no direct substitution between the inputs in various micro-units. Firms can, however, choose between different types of capital equipment and in this way obtain different factor proportions (or input coefficients). This dispersion of technology between plants makes substitution at the industry level possible, since a given amount of output can be produced with different combinations of plants. Therefore the substitution properties at different points of the isoquant map vary a lot.

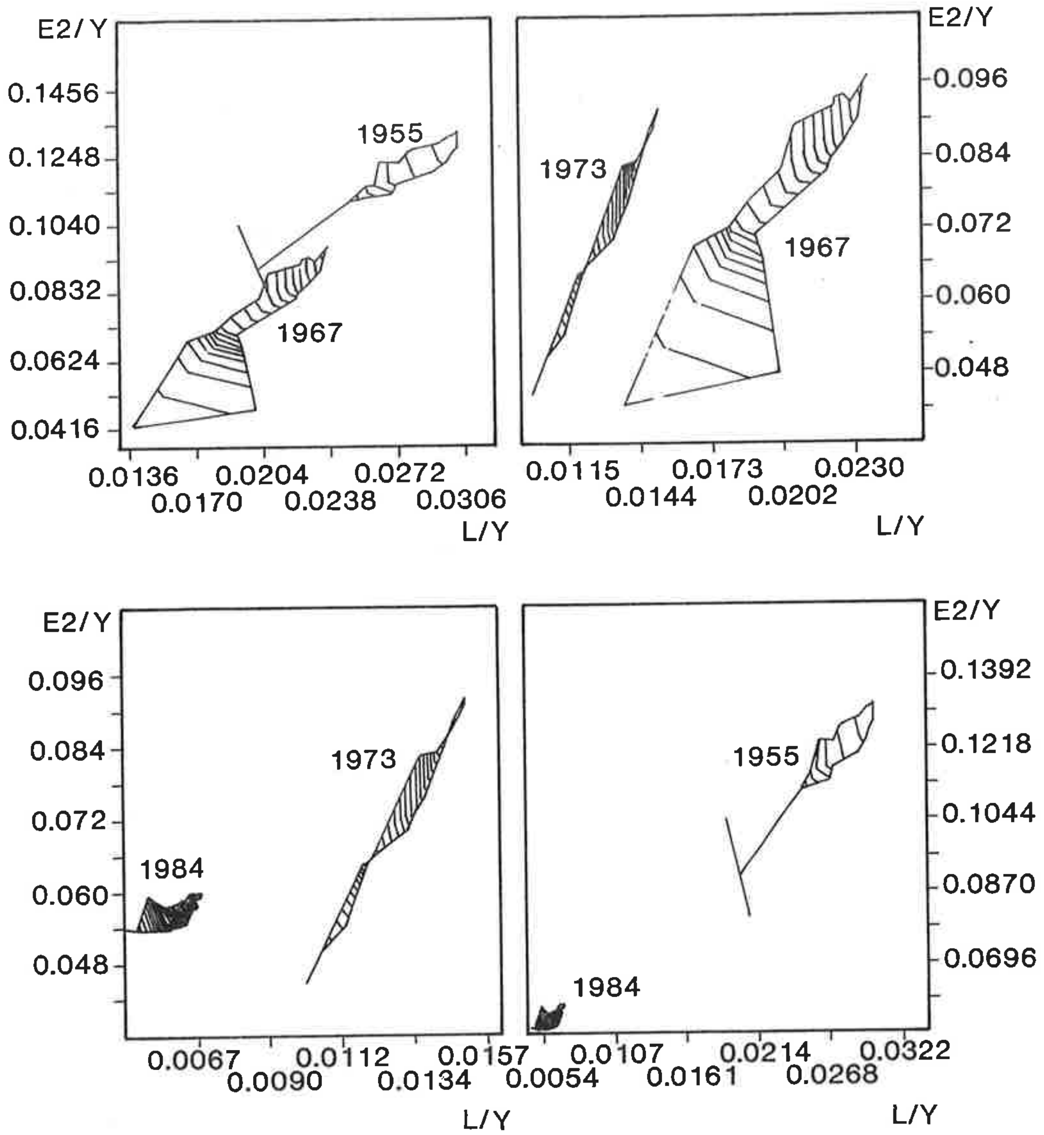
The slopes of the isoquants change considerably from year to year and even in the same year at different production levels (Figures 3.14 and 3.16). For instance in 1967 there is relatively little scope for labour substitution along the isoquants of the least efficient capacity, but when the most efficient part of the capacity is used, the isoquants take on a gentle slope and show more room for labour substitution. In 1973 most of the capacity is very inelastic in the labour input dimension (the isoquants are almost vertical), but more flexible in the energy input dimension.

The capacity region indicates the range within which the coefficients of two inputs, e.g. labour and energy, may fall when we are examining the SRIPF (Figure 3.16). An industry's capacity region also shows how its input coefficients vary with fluctuations in the relative input prices. Each subperiod, however, shows its own characteristic features, apparent in the typical changes occurring in the shape of the capacity regions on the one hand and in their location in the input coefficient space on the other.

It is difficult to graphically show all the capacity regions of the different years in one figure, because the development in 1955-1980 was too fast. During the observation period the capacity regions moved markedly towards the origin, i.e. productivity, in general, grew considerably. In Figure 3.15 we also look at the changes during each subperiod, all differing from each other in the character of their development.

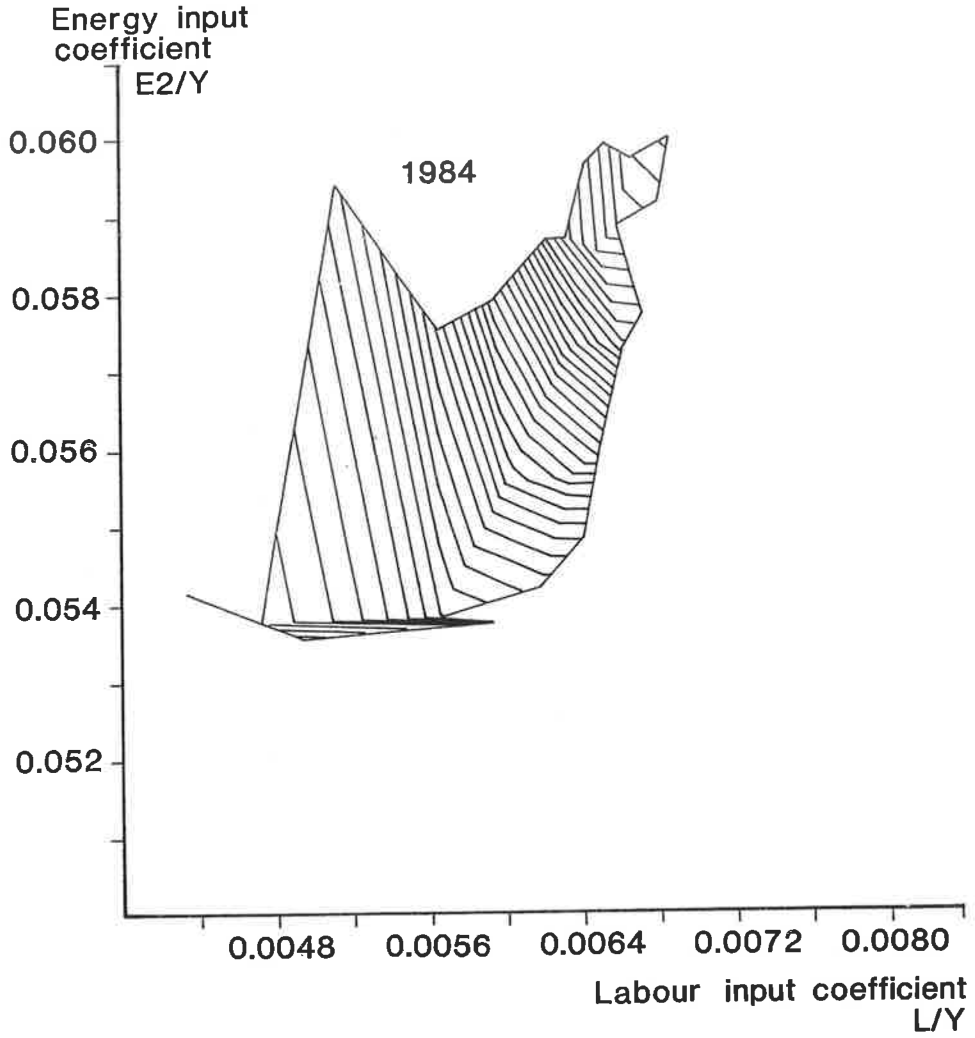
In 1955 the frontier capacity consisted of two plants; these, however, showed distinct differences in the energy-labour input dimension, which formed a straight line northwest-southeast. Aggregated, these plants form a southwest-northeast straight line, on which the capacity utilisation of the less efficient plants is growing. Only when the total capacity of the industry exceeds the aggregated capacity of these two plants do other plants enter the picture and the capacity region begins to show typical characteristics of a capacity region normally made up of heterogeneous units (cf. Figure 3.16).

Figure 3.15: Development of capacity regions in the labour - total energy space during the whole observation period and by subperiods.



In 1967 variations in the labour input coefficient at low output levels are much larger than for energy; at a higher utilisation levels the situation is reversed, as was the case in 1955. The highest substitution elasticity is found in the region where the energy/labour price ratio is low. At a lower capacity level the highest substitution elasticity is found in a region where the energy/labour price ratio is high.

Figure 3.16: The capacity region in 1984. The isoquants correspond to those of the short-run industry production function. The distance between the isoquants is 10 000 m^3.



In Figure 3.16 we also present a more detailed figure of the capacity region in 1984. At relatively low output levels the isoquants have a marked kink gradually decreasing towards higher output levels though returning on some isoquants at the top again.

3.4.3. Cost functions

We proceed with the analysis, this time based on the Heckscher diagrams and cost functions. The marginal cost curve in Figure 3.18 is defined as the increment in total variable costs, brought about by the employment of a new unit along the expansion path, divided by the corresponding increment in total output. Along the expansion path, only the capacity of one unit at a time is expanded to full capacity. The micro foundation of the marginal and average cost functions is presented in the form of the Heckscher diagrams, with unit costs on the ordinate axis and percentage capacity shares of the individual plants on the abscissa axis. In Figure 3.17 the costs are divided into unit wage costs and unit energy costs. We use the same input prices for all units due to the facts discussed earlier in Section 3.1. Thus the cost differences depend on the input coefficients only.

There is a close relationship between the Heckscher diagrams in Figure 3.17 and the marginal and average cost curves in Figure 3.18. Along the expansion path the marginal cost function is derived by expanding one production unit after another in the order of the ranking given by the corresponding Heckscher diagram. This transformation is performed by moving from the percentage output share-unit cost space into the output-cost space. The average cost curve is obtained by accumulating costs in the Hecksher diagram and weighting them by output or capacity shares. These relationships may give us a deeper understanding of the derived marginal and average cost functions of the industry.

The shapes of the marginal and average variable cost functions flattened out considerably when we compare the years 1955 and 1967 and the post investment boom period, with the exception of the least cost-efficient units (the right end of the curve) in 1973 and 1984. Firms adapted themselves to the wide changes through extension of marketing areas, major expansion of capacities and changes in the relative prices. Large differences in cost structure between plants up till the last years of the observation period have, nevertheless, been maintained. It is not until 1984 that a levelling out curve indicates major adaptation in the form of very slight cost differences.

It is especially interesting regarding the subperiod 1955-1967 to compare the results derived from the SRIPF on the one hand and from the marginal and average variable cost functions on the other. The SRIPF yields the impression of a static development, but the cost functions show significant improvement in the cost efficiency due to the neutral character of technical progress.

The Heckscher diagrams in Figure 3.17 give the wage and energy unit costs at 1984 prices. It is worth noting that the unit cost scale at the ordinate axes is different each from year to year. Unit costs show a clear decrease in all the successive years of comparison. The share of energy costs decreases from 1955 to 1967, but increases slightly during the investment period due to new, energy intensive technologies. The share of energy costs continues to increase in the mid and late 1970s, but now the higher energy prices explain the development. It is only towards the end of the observation period that the energy saving measures adopted by the firms make themselves felt - the share of energy costs stops increasing. The differences between the most and least cost-efficient units steadily decrease during the observation period, with the smaller units tending to concentrate at the least cost-efficient end of the diagram.

Figure 3.17: Heckscher diagrams, unit costs FIM/litre at 1984 prices.

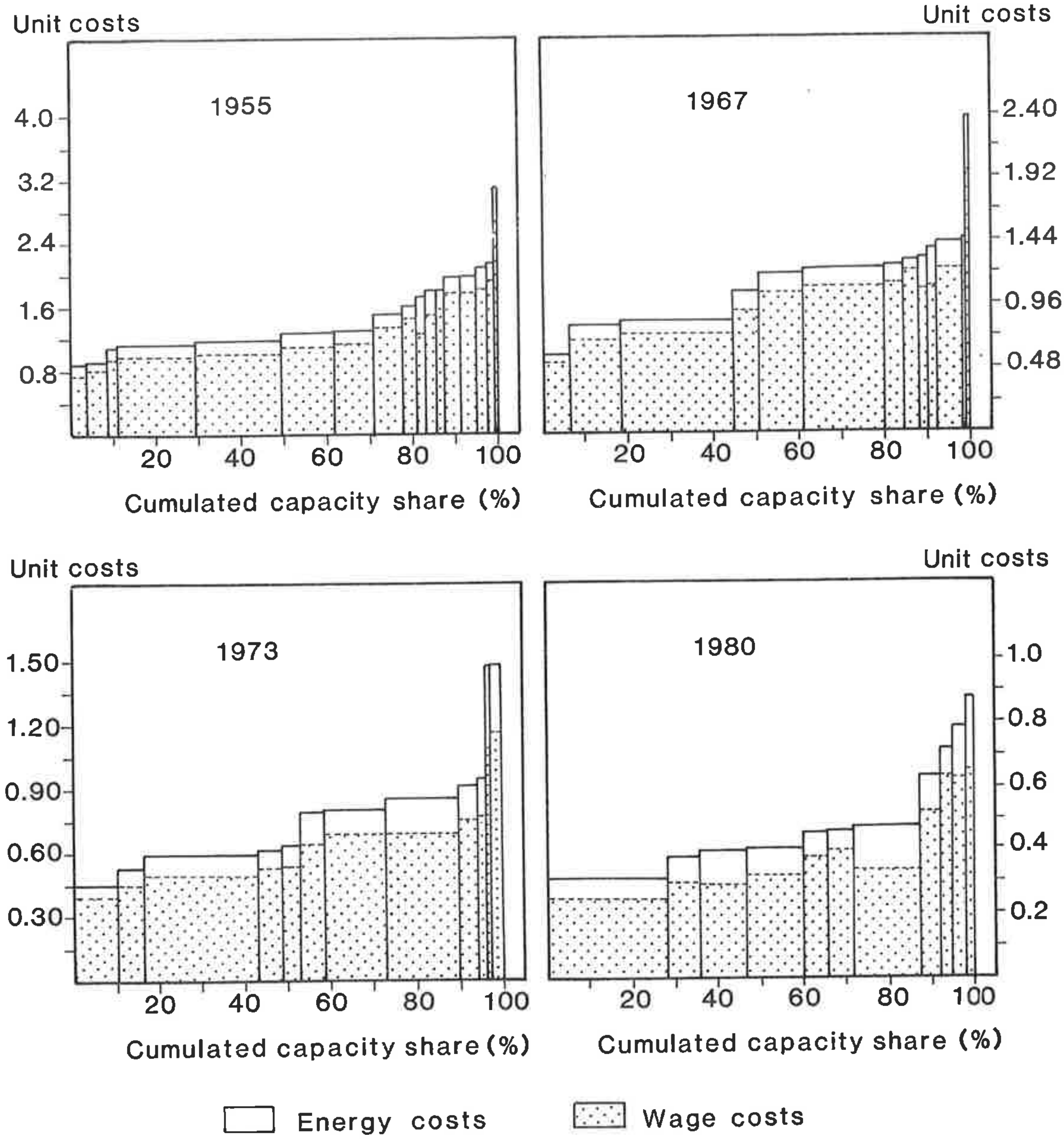
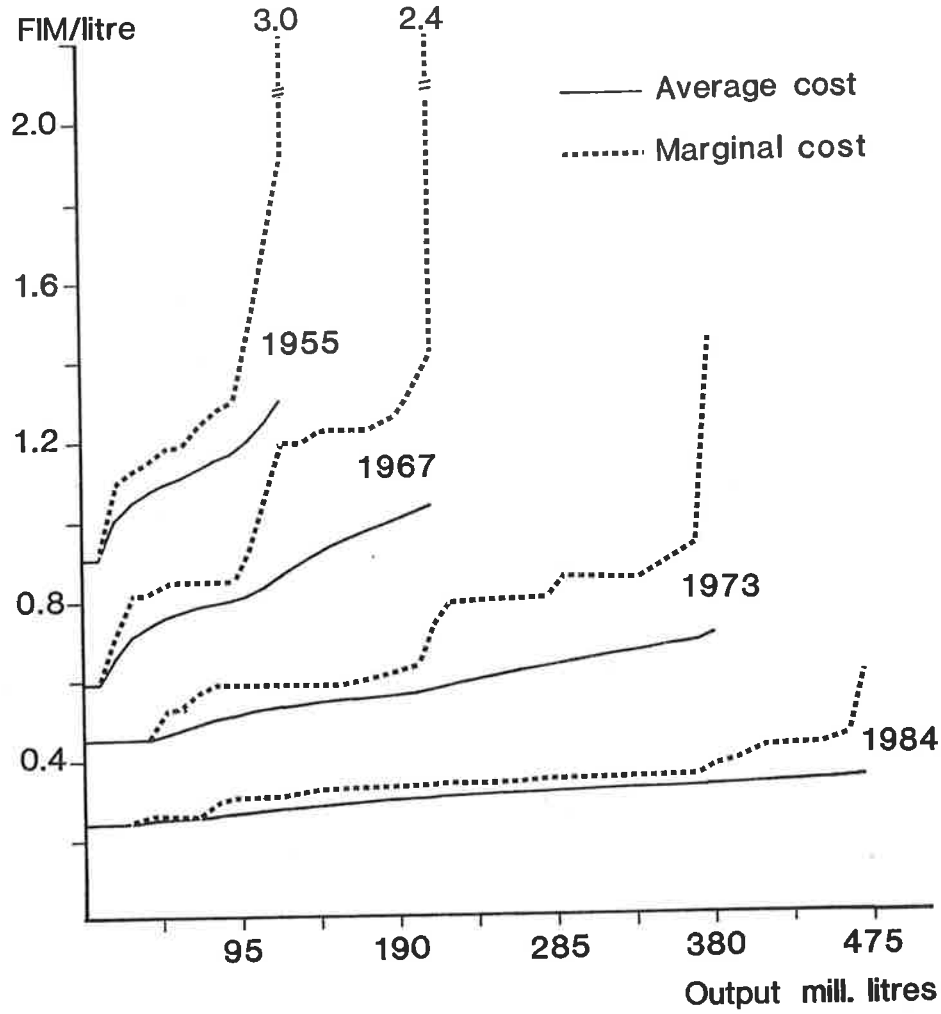


Figure 3.18: The marginal and average cost functions in 1955, 1967, 1973 and 1984 at 1984 prices.



3.4.4. Technical advance and bias

To get a more precise measure of technical change we shall calculate the Salter measures of technical advance and factor bias introduced in Section 2.3.3. We have chosen to use 1984 prices (Paasche-Konüs index) and have calculated the degree of technical progress and the factor bias for a few output levels in addition to the frontier of the capacity region shown in Figure 3.15. The empirical results are presented in Tables 3.14 and 3.15.

As can be seen from the tables, the rate of technical change has varied a lot. Between 1955 and 1967 it was very uniform, amounting to a yearly reduction of 3.1 % in unit costs at all output levels. Between 1967 and 1973 technical progress increased considerably, yielding an annual unit cost reduction of up to 10 % at the highest output level. The current unit cost reduction was also rapid between 1973 and 1984, around 6 % yearly and increasing somewhat when we move from the frontier to higher output levels. Thus this way of measuring technical advance confirms and quantifies the impression we received from Figure 3.15, i.e. that technical progress was very rapid not only between 1967 and 1973 but also after the heavy investment period as well. The results clearly confirm the importance of a fine-tuning phase in a process industry after new machinery and technologies have been applied. In general, between all the comparison years in Table 3.14 we get smaller cost reductions at lower output levels than at higher ones.

In Table 3.15 the bias measures are presented. The factor bias measures generally show a strong labour saving - energy using bias, except for the years 1955 to 1967. As mentioned above, this may be surprising because the relative price of labour increased by about 60 % during this period. The rapid capacity expansion between 1967 and 1973 markedly increases the optimal energy/labour ratio. Nevertheless, it is surprising to see that the same development continues even after 1973 and at a higher degree. After 1967 energy using bias is in general the stronger the lower the aggregate output level is and the optimal change in the energy/labour ratio exceeds 9 % at the frontier between 1973 and 1984, in spite of a small increase in the relative price of energy during this period.

Table 3.14: The Salter technical advance measure T at 1984 prices. $T = C_{t_1}/C_{t_0}$, where C_t = minimised unit cost at $y=y^0$ in year t. Annual percentage changes in parentheses.

Year	Output levels, y, in 1000 m³							
	Frontier	50	100	150	200	250	300	350
1967/1955	0.67 (3.1)	0.69 (3.1)	0.68 (3.1)					
1973/1967	0.76 (4.5)	0.62 (8.0)	0.64 (7.4)	0.58 (9.1)	0.55 (10.0)			
1984/1973	0.53 (5.8)	0.54 (5.6)	0.52 (5.9)	0.53 (5.8)	0.53 (5.8)	0.51 (6.1)	0.49 (6.5)	0.48 (6.7)
1984/1955	0.27 (4.5)	0.23 (5.1)	0.23 (5.1)					

Table 3.15: Factor bias. Change in optimal energy/labour factor ratio $(E2/L)_{t_1}/(E2/L)_{t_0}$ at 1984 prices. Annual percentage changes in parentheses.

Year	Output levels, y, in 1000 m³							
	Frontier	50	100	150	200	250	300	350
1967/1955	0.57 (-4.7)	0.91 (-0.8)	0.92 (-0.7)					
1973/1967	1.43 (6.0)	1.12 (1.9)	1.29 (4.2)	1.35 5.0)	1.35 (5.0)			
1984/1973	2.82 (9.4)	2.53 (8.4)	2.24 (7.3)	2.03 (6.4)	1.88 (5.7)	1.80 (5.3)	1.73 (5.0)	1.63 (4.4)
1984/1955	2.29 (2.9)	2.59 (3.3)	2.64 (3.3)					

Since we should expect the new capacity to be represented at the frontier, it is particularly interesting to look at the development of the ex ante production function during the course of the different subperiods. Our frontier production function may be regarded as a function representing the ex ante (or at least the best-practice) technology during the different periods. Returning to Table 3.14, it turns out that technical change in the frontier function was labour-using and energy-saving in 1955-1967 and labour-saving but energy-using in 1967-1973 and 1973-1984.

According to Table 3.15 for constant factor prices, the frontier production function yields an optimal decrease in the energy/labour ratio at constant factor prices of about 60 % in 1955-1967, which is about the same level as that obtained at the frontier in the SRIPF. At higher output levels, the influence of relative price changes should be less evident, which is consistent with a much smaller bias at higher output levels in the SRIPF. However, in 1973-1983 the energy-using bias is considerably higher at low output levels in the SRIPF than in the frontier function, in spite of an almost constant relative price ratio. The reason for this might be the presence of a disembodied labour-saving technical progress, as indicated by the movement of the capacity region in Figure 3.15. This period should be more strongly influenced by disembodied technical progress through fine-tuning of earlier investments than other periods are. It might also be the result of a lag effect due to strong labour-saving bias in the ex ante technology of 1967-1973.

In 1967-1973 the frontier function shows a very strong labour-saving and energy-using bias. We should bear in mind, though, that this bias is obtained as the difference between the marginal elasticities of two different frontier functions for the two subperiods respectively, so the results must be interpreted with some care. Moreover, the relative price of labour increases considerably. Against this background one might perhaps expect a more pronounced labour-saving bias also in the SRIPF.

4. STRUCTURAL EFFICIENCY AT THE INDUSTRY LEVEL

4.1. Measures based on the frontier production function

Since production units included in the SRIPF were once chosen from a menu of production possibilities represented by the frontier production function, there should be a connection between the development of the frontier function and the SRIPF. This does not mean, however, that we should expect the same type of technical progress or rate of technical advance in the two functions, since the frontier function traces the technological possibilities on the basis of the most efficient units, while the SRIPF is based on the whole sample. Moreover, the short-run industry production function is limited to current inputs, while capital is a variable in the frontier function. The development of the frontier function indicates that there are possibilities for productivity improvements for <u>one</u> unit, while the SRIPF shows the productivity improvements for the entire industry.

In this chapter we consider the structural efficiency at the industry level by using the information derived both from the frontier production function and from the SRIPF.

According to Farrell (1957), the purpose of a structural efficiency measure is to measure "the extent to which an industry keeps up with the performance of its own best firms". The approach, looking at the aggregated picture of the industry, suggested by Farrell, is to weight the individual measures by observed output levels.

Above, in Section 2.3.1. we introduced the efficiency measures E_1 (potential input saving), E_2 (potential output increasing) and E_3, E_4

and E_5 (potential reduction in input coefficients by producing at optimal scale) and presented the empirical results in Section 3.3.4. Now we introduce the structural measures reflecting the same properties for the industry as the E-measures for a micro unit.

The first measure of structural efficiency suggested by Farrell and here denoted by S_0 is obtained by taking the average of E_1 technical efficiency measures with outputs as weights. This weighting scheme has no straightforward interpretation in terms of the objectives of the structural measures, i.e. in terms of resource saving or output increasing. However, the reason for calculating S_0 is that it seems to be the only measure of structural efficiency that was used before Førsund and Hjalmarsson (1979) extended the Farrell analysis on this point.[1)]

In order to get a more satisfactory structural efficiency measure, something that could be explicitly interpreted in terms of input saving or output augmenting for the industry, Førsund and Hjalmarsson (1979) constructed an average plant for the industry and regarded this average plant as any other observation and then computed E_1 - E_5 for this average unit.[2)] These measures of structural efficiency are denoted by S_1 - S_5, where S_1 and S_2 are measures of structural technical efficiency, S_3 is a measure of structural scale efficiency and S_4 and S_5 pure structural scale efficiency measures.

1) See e.g. Carlsson (1972).

2) In this study the average plant was constructed by taking the arithmetic average of inputs and outputs.

For the last two measures it is possible to show analogously, as we did earlier in Section 2.3.1 for E_4 and E_5, that

$$S_4 = \frac{S_3}{S_1}$$

and that

$$S_5 = \frac{S_3}{S_2}$$

In this case there exists a clear relationship, shown in Førsund and Hjalmarsson (1984), between the scale properties of the production function and the efficiency measures. Since the average unit can be regarded as an arbitrary observation, the relationship between the different measures of structural efficiency and the average of the elasticity of scale is the same as the relationship between the corresponding E_1 measures. Thus

$$\bar{\varepsilon} = \frac{\ln S_2}{\ln S_1}$$

and

$$\bar{\varepsilon} = \frac{\ln S_3 - \ln S_5}{\ln S_3 - \ln S_4}$$

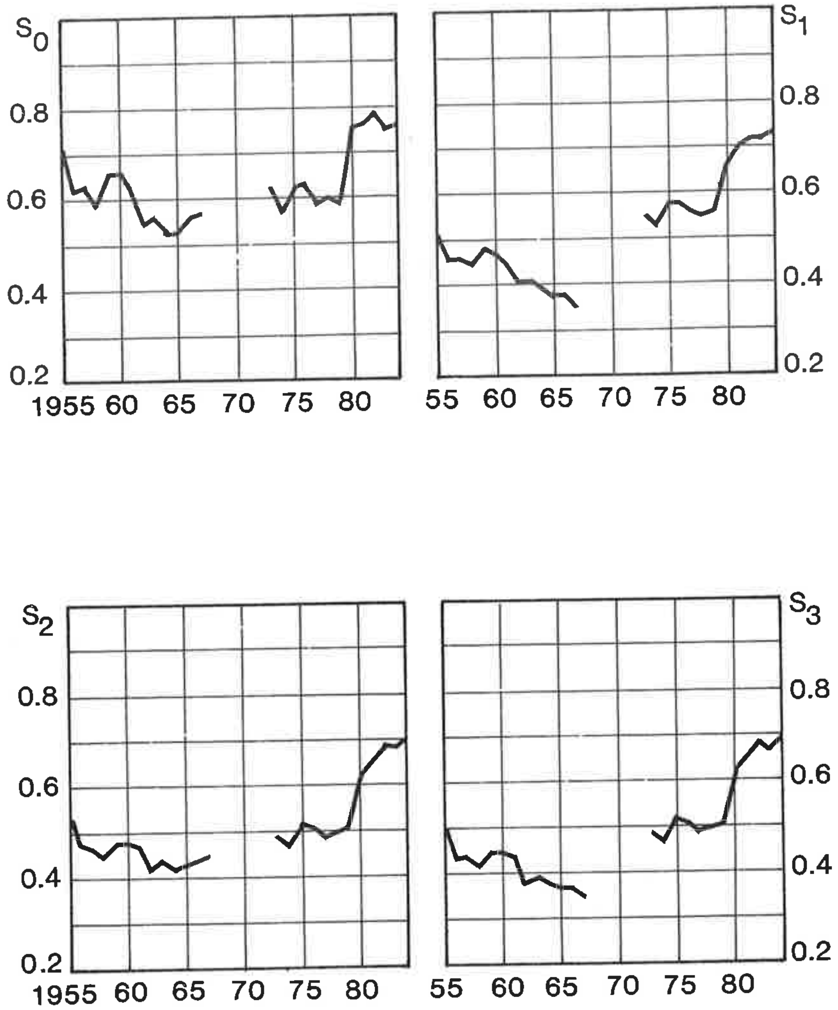
Because of the analogy with the E_1-measures, S_3 always shows a lower value than S_1 or S_2, except in the case where the industry consists of a number of plants of optimal size employing the same best-practice technique, a situation characterising a long run equilibrium of an industry; i.e. when all units produce at optimal scale.

In Figure 4.1 three different measures, in addition to S_0, are presented and in Appendix 3 we have included also S_4 and S_5. S_1 measures the distance between the average plant and the frontier at the observed average level of output. The measure shows the relative reduction in the amount of inputs needed to produce the observed average output, using frontier function technology with the observed factor proportions (input saving measure). The output increasing measure, S_2, is the ratio between the average observed output and the output obtainable on the frontier function using the observed average amount of inputs. The first measure of scale efficiency, S_3, shows the distance in terms of input coefficient reduction, from the observed average plant to the optimal scale at the frontier function, while the last two scale efficiency measures show the distance of the average plant from the optimal scale after moving the units to the frontier horizontally, S_4, and vertically, S_5, thus eliminating the two types of technical inefficiency before the scale efficiency is calculated.

The values of S_0 turn out to be the highest ones. They show a decreasing trend during the static period 1955-1967 but an increasing trend between 1973 and 1984. The values begin at 0.73 in 1955 and end at 0.57 in 1967. After the investment phase the starting values are somewhat higher, staying at about 0.60 during the years of fine-tuning and then rising to 0.75 and higher in the 1980s.

In his analysis of 26 Swedish industries Carlsson's (1972) S_0 estimate for Swedish breweries in 1968 was 0.76, among the least efficient ones. According to Carlsson's results, the least efficient industries are protected from foreign competition and the three most efficient

Figure 4.1: Development of S_0, S_1, S_2 and S_3 measures of structural efficiency.



industries are among the chief exporters in the Swedish manufacturing. On the average, our S_0 structural efficiency estimate in 1967 was even lower than that of Carlsson which may be due to a lot of facts. Methodological differences may be one reason. Carlsson estimated a Cobb-Douglas LP-frontier for a single cross section. Though the fit of a Cobb-Douglas function may be worse than for a homothetic function a single cross section should yield higher estimates of efficiency.

Another reason may be the more steady structural change and sharper market competition in Sweden. The value of our S_0 was for most years also lower than that of Swedish dairies, estimated to be in the range of 0.61 to 0.78 by Førsund and Hjalmarsson for the period 1964-1973. Dairies had the second lowest ranking in Carlsson's study. It is worth noting that with longlived equipment, as in our case, one should expect a wide dispersion between the input requirements of different units, i.e. one should expect a relatively low value of structural efficiency.

It can be seen clearly in Figure 4.1 that structural efficiency, as measured by S_1, S_2 and S_3, gives a picture rather similar to the development of S_0 though the levels differ. The higher efficiency noted for 1973-1984 shows that the industry succeeded in making good use of the large investments completed in 1967-1972. The jump following the investments in all the structural efficiency estimates confirms that technical progress is embodied but also that it takes some years before new equipment reaches its potential efficiency. We are thus faced with a mixed development emphasising embodiment at the first stage and at a latter stage disembodiment caused by learning-by-doing effects etc.

In our case there also are some interesting results regarding pure scale efficiency structural measures (see Appendix 3). The S_4 measure shows that the industry made effective use of the inputs in 1955-1967 - the estimated values ranging from 0.94 to 0.99 - but was not as succesful in 1973-1984 when values varied between 0.88 and 0.95. The S_5 completes the picture by showing the opposite results: during the years 1955-1967 the average output was not high enough (most years the S_5 values varied from 0.78 to 0.94) but after 1973 the average output was at about the scale optimum (S_5 gets a value of 0.99 or 1.00 in all years).

4.2. Changes in the price of production cost

When calculating Salter's technical advance measure for the short-run industry production function above we obtained the change in unit costs of production between t_0 and t_1 at a given output level for the prices of a certain year. However, it may also be interesting to compare unit costs of production at t_0 with those of t_1 for the same output level but at each year's factor prices. Then we obtain true price indices of production cost.

The formulas for the calculation of true price indices of production cost presume, however, a somewhat different selection of prices at which the Salter measures are to be calculated by subperiod, in order to achieve a consistent measure of price changes by subperiod to the price change over the whole period. Consistent aggregation over time must at the same time apply to the true indices and also to the measures of technical change and the corresponding price standards.

From the basic formulae (3) in section 2.4.4. it may be deducted that consistent aggregation over time may be done in the following way (cf. Karko (1986)):

$$\frac{C^T(y^*,\ w^T)}{C^0(y^*,\ w^0)} = \prod_{t=0}^{T-1} \frac{C^{t+1}(y^*,\ w^{t+1})}{C^t(y^*,\ w^t)} \tag{1}$$

$$= \prod_{t=0}^{T-1} \frac{C^{t+1}(y^*,\ w^{t+1})}{C^t(y^*,\ w^{t+1})} \ \frac{C^t(y^*,\ w^{t+1})}{C^t(y^*,\ w^t)}$$

The true price index of production cost over the whole period (0,T) is here multiplicatively decomposed into two components at the chosen output level y*. The first component consists of the Salter measure of technical change by subperiod $(t,t+1) \in (0,T)$, $t = 0,1,2,\ldots,T-1$. The formula applied here represents the Salter measure in the Paasche form by subperiod. This means that the measure is calculated at a certain chosen production level y* but at prices prevailing in the end years of each subperiod. The price base of the Salter index is thus a moving one as are the technical bases of the Salter indices.

In calculating the second component, the price standard or economic price index, the technology of the first years of each subperiod must serve as a base. This selection guarantees consistent chaining of the individual indices over the subperiods, to show an overall price change of the production cost. The procedure proves that calculating the indices by the base method, i.e. directly over the whole period, leads to the same result as calculating the indices via chaining the corresponding indices of the relevant subperiods. Moreover, this method also allows moving chaining, leading to the same result of relative change as the corresponding base method.

Table 4.1: The true price index of production cost (A), Salter measure of technical change in the Paasche form (B) and economic price index of production cost calculated at the technology of the first year of the (sub)period (C). Output level 100.000 m^3.

	(A)	(B)	(C)
1967/55	1.57	0.67	2.34
1973/67	1.33	0.66	2.01
1984/73	2.45	0.52	4.72
1984/55	5.10	0.23	22.19

From Table 4.1 we see, firstly, that there are no significant differences between the rates of technical change by subperiod calculated at the observed prices of the end years of the subperiods and the rates calculated at 1984 prices at the same output level (Table 3.7). The Salter measures are, naturally, the same over the entire period, having been calculated at 1984 prices for both tables.

The economic price indices by subperiods show how much more it would cost to produce a 100,000 m^3 output at the prices of the end-year of each subperiod w^{t+1}, compared to the cost that would be incurred at the prices of the first year of the corresponding subperiod w^t, by applying the first-year technology. The economic indices fluctuate widely (even in annual average terms) by subperiod. In this connection, the meaning of economic price index must also be emphasised. It is fully hypothetical for its constant technology. For instance, the economic price index over the whole period states that if the production technology in 1984 is identical with that of 1955, the cost of producing one m^3 at the output level of 100,000 m^3 for the whole industry in 1984 at 1984 prices would be 22 times as high as the cost of producing the same unit at the same industrial output level at 1955 prices.

Although the price of energy in the brewing industry has rosen 6.5-fold and the price of labour 10.5-fold, the true price of production cost at this level of output has only risen about 5-fold. This means that the true price of producing one unit in 1984 by using the 1984 technology has risen by 400 %, compared to the cost of producing one unit with the 1955 technology at the 1955 prices. Technical change has, thus, played a significant role in diminishing the price of production cost via reduction of factor demands on the same isoquant levels.

If we apply the Salter measures calculated on the basis of 1984 prices on a certain, identical output level in each subperiod, we may aggregate the Salter measures directly over the subperiods according to the following formula:

$$\frac{C^T(y^*, w^T)}{C^0(y^*, w^T)} = \prod_{t=0}^{T-1} \frac{C^{t+1}(y^*, w^T)}{C^t(y^*, w^T)} \qquad (2)$$

On the other hand, we may compute the rate of technical change over the whole period at the same prices also implicitly, without knowing explicitly the rates of technical change prevailing at each individual subperiod but knowing the corresponding true and economic price indices, by using the following extension of the deflation procedure:

$$\frac{C^T(y^*, w^T)}{C^0(y^*, w^T)} = \prod_{t=0}^{T-1} \frac{C^{t+1}(y^*, w^T)}{C^t(y^*, w^t)} \Big/ \prod_{t=0}^{T-1} \frac{C^t(y^*, w^T)}{C^t(y^*, w^t)} \qquad (3)$$

Thus the Salter rate over the whole period may be obtained by dividing the product of the true price indices of the subperiods by the product of the corresponding economic price indices. Each true subindex is calculated as a ratio of the end- and first year technology of the relevant subperiod. In the true and economic index, however, the price base is the price vector of the first year of each subperiod; in the cost functions of their numerators, however, the price vector is that of the end year of the entire period. It should also be noted that this method allows us to gradually enlarge the observation period by adding more subperiods.

Finally we check the validity of the methods that we have applied on our data and construct the following table:

Table 4.2: The explicit and implicit rates of technical change at 1984 prices by gradual extension of subperiods

	True price index		Economic price index		Rate of technical change (Paasche form)	
	By sub-period	Cumu-lative	By sub-period	Cumu-lative	By sub-period	Cumu-lative
1967/55	2.45		4.72		0.52	
1973/67	6.26		9.78		0.64	
1973/55		15.35		46.22		0.33
1984/73	15.36		22.19		0.68	
1984/55		235.58		1025.38		0.23

The prices at which the Salter measures are to be computed need not be those of the end year of the entire period, any comparable fixed prices may be used. The same is also true of the selection of the base prices. We only must see to it that the prices are selected systematically in such a way that formulae (2) and (3) are maintained. On the other hand, we also must, naturally, treat the technology in the index formulae systematically. The technologies depend on the subperiods considered and are independent of price selection, because production functions contain all possible prices. It may be possible to get different rates of all the indices used here by using different prices as well as different output levels.

5. CONCLUSIONS AND EXTENSIONS OF THE PRESENT STUDY

Empirical observations show us that not all firms operate with the same degree of technical and economic efficiency. We have studied efficiency on the micro and industry level and its linkage with industrial efficiency structure within the Johansen-Salter framework. The role of technical progress, one of the key factors influencing the efficiency, has been emphasised.

In an industry there often is a distinct difference between the substitution possibilities before and after the actual construction of plants. This aspect is most clearly captured by the putty-clay approach assuming smooth substitution possibilities ex ante and fixed coefficients for current inputs and capacity determined by the initial investment ex post. We have used the production function approach introduced by Johansen (1972), integrating the above-mentioned properties.

Within the framework applied it is necessary, at the micro level, to distinguish between the production possibilities existing before the time of investment - the ex ante or frontier production function - and those existing after the investment - the ex post production function. Aggregating the ex post functions of the micro units, at a certain point of time, yields the short-run industry production function.

The main contribution of this study is an empirical analysis of technical progress in an industry based on these two production function concepts simultaneously. Both deterministic and stochastic models have been used in estimating the frontier production function. The impact of technical advance has been measured in sevaral ways, using generalisations of Salter's and Farrell's approaches. New measures for decomposing the efficiency at the industry level are applied.

The data used is not only of high quality but also exceptionally well suited for the testing of the analysing tools, because the observation period is long and includes three different phases: the phase of relatively slow (about 4 per cent per year) growth in 1955-1967, the phase of heavy investments and radical production volume growth (about 10 per cent per year) in 1968-1972 and the phase consisting of the years 1973-1984 when production growth was quite slow (about 2 per cent per year) but steep changes occurred in input prices. Another interesting feature is the gradual sharpening of competition in the course of the period.

On the estimation side, this study shows the advantage of using the deterministic frontier approach in studying the process of technical change in an industry when we have access to reliable data. Our results illustrate fairly well the main weakness of the stochastic frontier approach: In the case of a fairly symmetric error distribution, the stochastic frontier is very close to the average function and a large share of productive inefficiency is erroneously explained by stochastic phenomena.

Since the production units comprised in the short-run industry production function once in the past have been chosen from the menu of production possibilities represented by the frontier production function, there should be a connection between the development of the frontier function and the short-run industry production function. However, this does not mean that one should expect the same type of technical progress or rate of technical advance in the two functions since the frontier function traces the technological possibilities on the basis of the most efficient units, while the short-run industry production function is based on the entire data. There are a lot of factors influencing the input coefficient of an individual plant, not included in a pure production function analysis, e.g. managerial efficiency, degree of competition, varying expectations concerning future prices and various random effects. Moreover, the short-run industry function is limited to current inputs, while in the frontier function capital is the variable. The development of the frontier function indicates the possibilities of productivity improvements for one unit, while the short-run industry production function shows the productivity improvements of the entire industry. Thus it is somewhat dangerous to draw any firm conclusions about the expected impact on the short-run industry production function of changes in the frontier function and relative prices, since capital is not included in the short-run function.

Table 5.1: A summary of the characteristics of technical change between 1955 and 1984[1)]

	Type of function	
	Frontier (deterministic)	Short-run
Bias		
1955/67	Weak energy saving	Weak energy using
1967/73	Strong labour saving	Moderate labour saving
1973/84	Moderate labour saving	Strong labour saving
Advance p.a.		
1955/67	4.9 %	3.1 %
1967/73	16 - 18 %	4.5 - 10 %
1973/84	0.9 %	5.6 - 6.7 %

1) See Tables 3.7, 3.9, 3.14 and 3.15.

We summarise the main results regarding the nature and rate of technical change in the industry studied in Table 5.1.

Our frontier production function may be regarded as a function representing the best-practice technology during the different periods. An interesting question is to which degree technical progress is embodied in new capital and to which degree choice of technology is governed by (expected) relative prices. If there is an embodied effect this should be evident only during a capacity expansion. We also observe a very strong impact on technical progress of the investment boom 1967-1973, a period during which the net stock of capital doubled. There was also a growth in the net stock of capital between 1955 and 1967, amounting to about 3 per cent per year.

During the last period, however, the net stock of capital decreased by 1.6 per cent annually. This is also reflected in a very small rate of technical progress in the frontier function.

The development of the short-run industry production function is also interesting. During the first period technical progress is relatively slow and almost neutral as regards input saving. On the other hand, during the boom years 1967-1973 there was a fairly strong, though not uniform labour saving bias. This bias continues even stronger during the last period. Since there was a strong increase in the relative labour/energy price up to 1973, we should expect this to have an impact on the choice of technology.

On the other hand, the relative labour/energy price decreased somewhat between 1973 and 1984. During this period, however, the capital stock decreased; we should, therefore, not expect any strong bias effect from the development of relative prices. Instead, we seem to observe a lag effect from the investment boom, resulting in a relatively strong rate of disembodied technical progress. It seems that the investment boom created a potential for future labour saving which took several years to realise.

It is also interesting to compare the differences between the frontier and short-run industry production functions in the degree of productivity slowdown after 1973. Even if there is a slowdown in both cases, it is much stronger in the frontier function, which indicates the importance of capital formation for productivity growth.

The analysing tools used in this study complement each other well, giving a detailed picture of the change of the frontier and the impact of this change on the character of technical change, in particular the impact on the change in unit cost at the optimal scale level. We have obtained the level and the development of productive efficiency both at the individual plant level and the entire industry level. The methods show great potential in industrial analysis.

This report is part of an ongoing work. There are several potential ways of extending the theoretical and empirical research. One new, interesting development of the stochastic frontier model, for instance, is to utilise panel data to obtain individual efficiency measures. The distance from the frontier due to inefficiency for each unit might be assumed constant over time. Due to the realisations of the pure random term, one would then not expect the ranking of the distances from the frontier to be systematic. Application and testing of the panel data model on this data set is currently under way.

Other relevant areas for further research are the use of more general functional forms (e.g. translog) and the study of relationships between the frontier and the short-run industry functions. Attempts to apply the methods to a more heterogenous sector, the engineering industry, is the subject of current research. Comparisons of industry data from several countries are also under way. The performance of the public sector is also an area of growing research interest.[1)] The progress made in industrial efficiency studies might be of use even in the strongly advancing research in the strategy of a firm.[2)]

1) For applications and theoretical research on the performance of public enterprises see Marchand, Pestieau and Tulkens (1984).

2) See e.g. Porter (1979) and (1980).

A P P E N D I X 1

DATA[1)]

The industry

The data was gathered from the Finnish brewing industry as defined in the Finnish industrial statistics: the ISIC sector 2132 in 1954-1970 and sector 31332 since 1971. The change in the statistical classification notwithstanding, there has been no change in the plants and products included in this industrial sector.

Data sources

The basic data was collected from each plant's annual Industrial Statistics information forms. These data were augmented, in respect of the construction of the capital variable of the years 1954-1958, by data obtained from the balance sheets of each firm. Since each plant, at that time, was an independent firm it was possible to augment the data this way. Special attention was paid to the quality of the data for which reason it was submitted to numerous check-ups in cooperation with representatives of the firms included and experts in the field.

Plants and time series

The data collected covered the years 1954-1984. Out of the original 29, 18 plants were accepted on the same criteria as were used by Airaksinen

---

1) Summa (1985) gives a deeper description of the data.

(1977) and Summa (1985). Data on these plants was available for different lengths of time, as follows:

10 plants for 31 years
1 plant for 24 years
1 plant for 16 years
1 plant for 14 years
3 plants for 13 years
1 plant for 11 years
1 plant for 9 years

The output of the plants included accounted for over 95 % of the output of the entire sector in all the years observed.

Operational variables

A. Output (Y)

In constructing the output volume variable, the production of each plant was first divided into two categories: malt beverages and soft drinks. This division can be unambiguously carried through product by product and it corresponds to the dual character of the production process.[1)] The litre volumes of the various kinds of drinks were added together, each within its own category. The soft drink volume was translated into beer equivalent and added to the volume of malt beverages to arrive at the total output in beer equivalent according to the following formula:

---

1) For more details see Summa (1985), pp. 21-22.

$$Y_{i,t} = YM_{i,t} + YS_{i,t} \cdot \frac{PS_{i,t}}{PM_{i,t}}$$

where Y = output volume in equivalent malt beverage litres

YM = output volume of malt beverages in litres

YS = output volume of soft drinks in litres

PM = average price of malt beverages, marks per litre

PS = average price of soft drinks, marks per litre

i = plants 1 - 18

t = years 1954-1984

PM and PS are calculated by dividing the value of the total quantity of malt beverages and soft drinks, respectively, produced by the equivalent production volume in litres. The method corresponds to the Leontief volume concept, which is strongly recommended e.g. in Vartia (1976).

The unit of measurement for the output volume (Y) is 1.000 litres = 1 m^3.

B. Labour (L)

The labour input chosen for this study is based on hours worked by all workers.[1] The labour input variable (L) is expressed in 1.000 hours.

The price of labour is obtained by dividing wages paid to all workers by L. The wage bill used is defined as in Industrial statistics and includes employers' contributions to social security. In defining prices

---

1) For other alternatives, see Summa (1985), pp. 23-24.

for all inputs we have used annual industry averages. Thus the prices do not differ by plants.

C. Capital (KT)

The capital input was constructed by using the so called bench-mark perpetual inventory method with fixed depreciation coefficient as in Airaksinen (1977) and Summa (1985).[1] The capital input series were constructed recursively using the equation

$$KT_{t_1} = INV_{t_0} + KT_t - D_{t_1}$$

where KT_{t_1} = current value of capital stock at year's end

INV_{t_1} = gross investment in the year t_1

D_{t_1} = depreciation in the year t_1

The current value of the capital stock in the basic year as a lower bench-mark was taken from the Industrial Statistics, for the new capital stock variable was constructed with the year 1954 as the starting point. Owing to the changes in the compilation of data for the Industrial Statistics, discussed in Summa (1985), the years after 1959 could not be accepted as the basic year. In 1954, the fire insurance values were first introduced in the Industrial Statistics, for which reason it might be safely supposed that both the Central Statistical Office and the firms paid special attention to the quality of the data. This is why we choose 1954 as the basic year for our calculations.

1) The alternatives and problems faced in constructing the capital variable are discussed in more detail in Summa (1985). The problem is to find out the depreciation coefficient ρ for $KT_{t-1}-D_t=(1-\rho)KT_{t-1}$.

The investment data have been treated as gross figures. Since fixed asset acquisitions were not introduced in the Industrial Statistics until 1959, the investment figures for 1954-1958 have been taken from the balance sheets of the firms. Each plant was an independent firm in those years, so that the data gathered this way correspond with the concept of investment applied by the Industrial Statistics later on.

The investment and capital stock data were deflated at 1980 prices by the subindex 7 of the official wholesale index for machinery and equipment and by the construction cost index for buildings.[1)]

No depreciation figures were available in the statistics so that we had to estimate them. We chose geometrical depreciation, giving us a constant depreciation coefficient for the entire period:

$$D_t = \rho KT_t$$

The depreciation coefficient ρ can be estimated on basis of the investment data plus the capital stock data for the first and last years or for two or three intermediary years of the observation period as bench-marks. The former procedure was chosen because we supposed that the data of the first year, the lower bench-mark 1954, in the material available were more reliable than the data for the rest of the years and because

---

1) There is an aggregation problem in using the price index of the latest vintages when the performance of the capital goods improves and the prices do not increase accordingly. So we tend to over-estimate the value of the capital stock. See e.g. Jorgenson and Griliches (1967) for the elimination of aggregation bias related to quality changes. This problem has recently been emphasised by Ylä-Liedenpohja and Törmä (1984).

we were able to check the data of most of the plants for the last year of the time series.[1)]

The unweighted average depreciation for machinery and equipment is 10.0 % and for buildings 4.2 %. These results coincide well with what experts consider an appropriate calculatory depreciation in this industry.

The series for all plants were calculated on basis of the same, average depreciation coefficients of the industry. The values of each plant in 1954 were used as starting values for all plants except for the one new plant where the year 1972 served as basis. The new series give a more realistic picture of the development of the net capital stock of the plants than do the fire insurance values in the Industrial Statistics; see Summa (1985).

The capital series were constructed separately for machinery and equipment and for buildings, but only the results based on the aggregated capital input variables are reported in this study.

The measurement unit for the total capital stock (KT) variable is 1.000 Finnish marks at 1980 prices.

The price of capital includes the average interest rate per annum for loans granted by commercial banks and the average depreciation rate. We have used the same depreciation rate for all years. Thus the price of capital varies yearly only due to the changes in the loan rates.

---

1) The basic principle of the computer programme used is explained in Summa (1985).

D. Electricity (EE)

Several potential ways of measuring the energy input were considered. The amount of electricity used can be measured with accuracy and the amount used specifically for production purposes determined more accurately than in the case of other forms of energy.

The electricity used (EE) data were obtained from the Industrial Statistics. The measurement unit is 1.000 kWh.

The price of electricity is obtained by dividing the value of electricity used by EE.

E. Energy (E2)

Electricity only covers part of the total energy consumption of the plants. During the observation period several kinds of energy were used in addition to electricity, e.g. fuels such as wood, coal, light and heavy oil, petrol and gas. In the course of the observation period, changes occurred in the volume shares of these various forms of energy used.

There are two main ways of constructing a variable for the total energy consumption:

(1) deflating the value of the total energy by the proper price index, or
(2) converting the different kinds of energy into equivalent energy units (e.g. joules or calories) and summing these up.

Since the quality of the volume energy data of some plants was poor, we decided to rely upon the former principle.

Firstly, we calculated the value of all kinds of energy used,other than electricity, and deflated the sum by the official energy price index[1], where the weights of the various kinds of energy used are fairly close to the average weights in the brewing industry. We expressed this series at 1980 prices.

Secondly, we multiplied the electricity used, expressed in kWh, by the average price of kWh paid by the industry in each year and expressed this at 1980 prices as well.

Thirdly, we added these two components together.

The total energy consumption (E2) is expressed in 1,000 Finnish marks at 1980 prices.

The subindex 31a of the official wholesale price index was used to indicate the price changes of the total energy consumption.

The industry total time series are represented in Figure 3.1 and the average annual values of the inputs are given in Table A.1. The relative prices between the inputs are found in Table 3.1.

---

1) Subindex 31a (mineral fuels) of the wholesale price index in the Statistical Yearbook of Finland published annually.

Table A.1: Average values of the variables used

Year	Period	Number of plants	Variable				
			Output Y 1000 litres	Labour L 1000 hours	Capital KT 1000 FIM*	Energy E2 1000 FIM*	Electricity EE 1000 kWh
1955	1	17	7587	230	13869	1004	497
1956	2	17	7365	244	14312	948	506
1957	3	17	7555	238	14340	944	537
1958	4	17	7327	233	13975	946	546
1959	5	17	8199	227	14602	939	618
1960	6	17	8765	229	15429	947	652
1961	7	17	9398	245	16078	1062	744
1962	8	17	8797	241	17600	1096	751
1963	9	16	9964	252	18373	1196	827
1964	10	16	10079	257	18442	1134	881
1965	11	15	11866	296	19584	1247	998
1966	12	15	13583	317	21140	1344	1095
1967	13	13	16663	394	27065	1622	1376
1968	14	12	19138	427	33938	1811	1646
1969	15	12	26550	520	38218	2193	2108
1970	16	11	30360	555	49010	2850	2714
1971	17	11	28793	518	48925	2739	2997
1972	18	12	31315	489	52089	2850	3159
1973	19	12	32011	485	61664	2940	3227
1974	20	12	32737	457	60604	3005	3077
1975	21	12	33172	430	58517	2760	3065
1976	22	12	31922	400	56339	2734	3276
1977	23	12	32570	399	57503	2683	3642
1978	24	11	36252	405	61914	2895	4030
1979	25	11	35697	375	59497	2862	3988
1980	26	11	41273	363	59468	2888	4323
1981	27	11	39735	346	58616	2673	4312
1982	28	11	40636	325	57318	2616	4266
1983	29	11	41092	312	56707	2609	4390
1984	30	11	43097	296	56355	2585	4307

* At 1980 prices.

APPENDIX 2

Spearman's rank correlation coefficient between different efficiency measures for selected years

	E_1	E_2	E_3	E_4	E_5
1955/56	0.94	0.97	0.96	0.97	0.98
56/57	0.94	0.91	0.98	0.97	0.96
57/58	0.84	0.96	0.94	0.96	0.99
58/59	0.93	0.94	0.93	0.99	0.97
59/60	0.89	0.91	0.96	0.98	0.97
60/61	0.97	0.92	0.98	0.99	0.96
61/62	0.87	0.95	0.90	0.96	0.93
62/63	0.96	0.92	0.92	0.97	0.97
63/64	0.89	0.87	0.93	0.99	0.97
64/65	0.93	0.79	0.92	0.98	0.94
65/66	0.95	0.77	0.87	0.96	0.96
66/67	0.82	0.94	0.89	0.96	0.96
73/74	0.83	0.78	0.94	0.97	0.67
74/75	0.77	0.85	0.84	0.97	0.87
75/76	0.94	0.88	0.84	0.99	0.90
76/77	0.62	0.64	0.75	0.98	0.89
77/78	0.82	0.90	0.91	0.99	0.99
78/79	0.86	0.93	0.95	0.98	0.87
79/80	0.66	0.83	0.88	0.92	0.94
80/81	0.86	0.92	0.95	0.96	0.86
81/82	0.85	0.85	0.95	0.88	0.91
82/83	0.55	0.47	0.59	0.69	0.99
83/84	0.89	0.56	0.67	0.98	0.96
1955/67	0.49	0.86	0.63	0.86	0.82
1967/73	0.33	0.28	0.19	0.15	-0.20
1973/84	0.08	-0.03	-0.03	0.62	-0.23
1955/84	0.13	0.38	-0.24	-0.22	-0.28

APPENDIX 3

MEASURES OF STRUCTURAL EFFICIENCY

Year	S_0 Weighted sum of efficiency measures.	S_1 The distance of the average plant to the frontier function for given output. (Corresponds to E_1)	S_2 The distance of the average plant to the frontier function for given amount of inputs. (Corresponds to E_2)	S_3 The distance of the average plant to the efficiency frontier. (Corresponds to E_3)	S_4 S_3/S_1 Pure scale efficiency. (Corresponds to E_4)	S_5 $S_5=S_3/S_2$ Pure scale efficiency. (Corresponds to E_5)
1955	0.73	0.51	0.53	0.50	0.97	0.94
56	0.62	0.46	0.48	0.44	0.96	0.93
57	0.63	0.46	0.47	0.44	0.96	0.93
58	0.59	0.45	0.45	0.42	0.94	0.94
59	0.66	0.48	0.48	0.45	0.96	0.94
1960	0.66	0.47	0.48	0.45	0.96	0.94
61	0.63	0.45	0.47	0.44	0.96	0.92
62	0.55	0.41	0.42	0.38	0.94	0.92
63	0.56	0.41	0.44	0.39	0.95	0.91
64	0.53	0.40	0.42	0.38	0.94	0.91
1965	0.53	0.38	0.43	0.37	0.97	0.87
66	0.56	0.38	0.44	0.37	0.98	0.85
67	0.57	0.35	0.45	0.35	1.00	0.78
73	0.62	0.55	0.49	0.49	0.88	1.00
74	0.57	0.53	0.47	0.47	0.89	1.00
1975	0.62	0.58	0.52	0.52	0.89	1.00
76	0.63	0.58	0.51	0.51	0.88	1.00
77	0.59	0.56	0.49	0.89	0.89	1.00
78	0.60	0.55	0.50	0.50	0.91	1.00
79	0.59	0.56	0.51	0.51	0.91	1.00
1980	0.75	0.66	0.62	0.62	0.94	1.00
81	0.76	0.70	0.65	0.65	0.93	0.99
82	0.78	0.72	0.68	0.68	0.94	0.99
83	0.75	0.72	0.68	0.67	0.94	0.99
84	0.76	0.73	0.70	0.69	0.95	1.00

APPENDIX 4

WORKING OF THE ALGORITHM: An Example[1)]

The purpose of this Appendix is to present by means of a numerical example the alogrithm for the computation of the short-run industry production function in more detail. The example refers to the 300,000 m^3 isoquant of the brewing industry in 1980, shown in Figure A.1. The basic idea of the algorithm is to compare angles between units in the input coefficient space[2)], i.e. the basis for the construction of isoquants is the slope matrix. This matrix consists of all angles between the production units in the input coefficient space. It is a triangular matrix without the main diagonal where the units are entered according to increasing input coefficients of the abscissa input, both along the rows and columns. The complete slope matrix is presented in Table A.2 (the rectangles around some figures are used in the construction of an isoquant). The boundary of the substitution region up to this isoquant level and the isoquant itself are presented in Table A.3.

The substitution region

The boundaries of the substitution region are determined by ranking the units according to increasing input coefficients for each input separately. This corresponds to sweeping horizontal and vertical "price" lines outwards from the axes over the capacity distribution (see Figure

---

1) The algorithm was developed by Førsund and Hjalmarsson and programmed at the University of Oslo.

2) A more detailed theoretical derivation of the algorithm is given in Førsund and Hjalmarsson (1984).

Figure A.1: The short-run industry production function in 1980. The distance between the isoquants is 10 000 m^3.

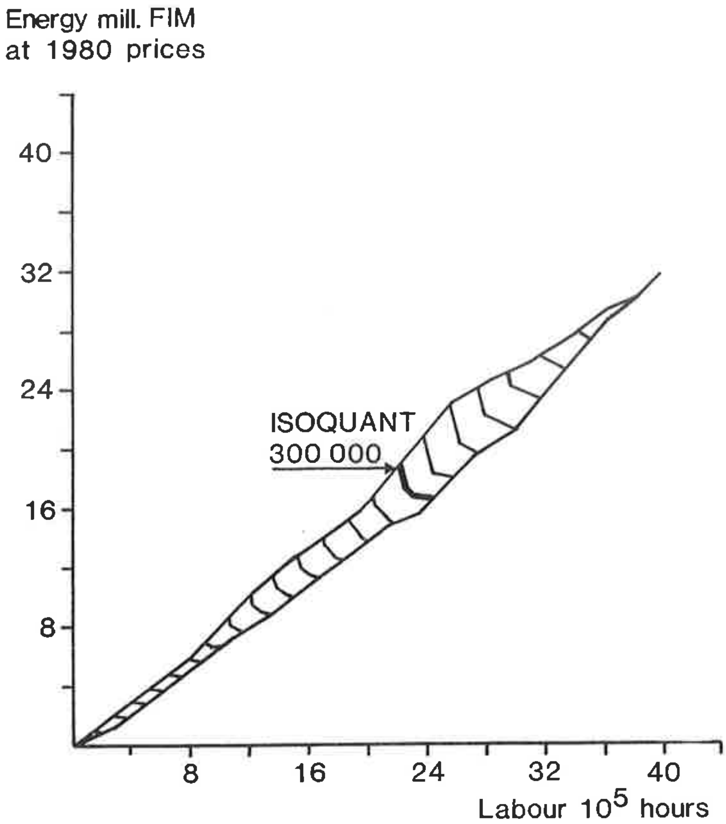


Table A.2: The S-matrix of slope coefficients in 1980, labour as abscissa variable.

Unit No	2	11	3	1	9	6	5	10	8	18
4	26.32	10.95	7.31	24.03	2.80	-0.23	5.08	7.59	1.63	12.21
2		-167.60	-22.04	21.58	[4] -9.72	-11.95	0.72	4.95	[5] -1.78	10.35
11			0.15	40.69	-2.14	-5.81	3.75	7.07	0.21	12.39
3				115.94	-3.16	-7.83	4.22	7.66	0.22	13.36
1					[1] -40.98	[2] -35.63	-4.27	2.43	-5.26	8.77
9						[3] -22.72	7.28	10.21	0.99	16.93
6							11.77	12.85	2.84	19.82
5								14.09	-6.88	27.46
10									-373.70	89.79
8										258.35

3.16) and entering the plant capacities in the order they appear. In Table A.2 (and A.3) the units are ranked according to increasing labour input coefficients (on the upper boundary).

This means that the slope matrix yields the order in which the units are utilised along the upper boundary of the substitution region. In our case the unit with the highest labour productivity is No. 4 which is utilised first along the upper boundary. The next units are No. 2, 11, 3, etc. according to Table A.2. The numbers of the units are marked off in Figure A.2.

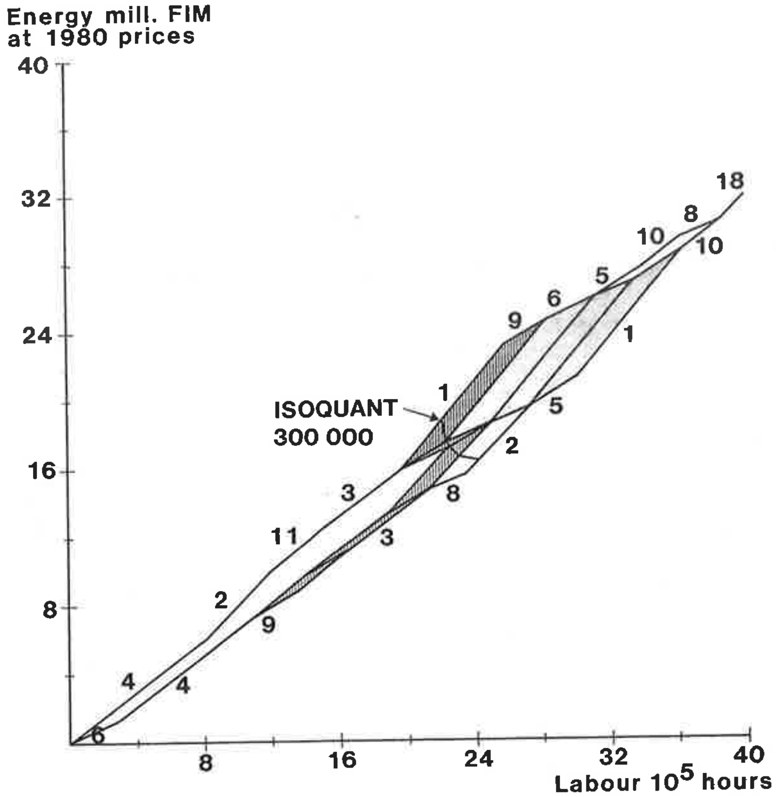
At the lower boundary we get another ranking, now according to energy productivity. This is also marked off in Figure A.2.

The Activity Regions

In a short-run industry production function the units of an industry are combined in the most efficient way. It might therefore also be useful to portray the complete set of efficient combinations of production units. This is done in Figure A.2. (For a more elaborate treatment of this see Førsund and Hjalmarsson (1984)). The pattern in which a production unit is used between 0 % and 100 % of the capacity is called its activity region.

Starting at zero industry production and expanding to full capacity utilisation,the activity regions are formed by adding micro units in accordance with the requirement that at each point in the substitution region, maximum industry output is obtained. Each parallelogram is formed by combining two units. Within the parallelogram the utilisation rate is between zero and one.

Figure A.2: The short-run industry production function in 1980. The isoquant 300 000 m^3 used as an example is marked with a heavy line.



This kind of representation gives us the possibility of following the utilisation of each unit as a function of the degree of capacity utilisation in the industry.

The units move in a strip-like fashion by parallelograms from one side of the substitution region to the other. South-west of the parallelogram the unit is not utilised while north-east of the parallelogram it is fully utilised.

The shape of the path of a specific unit is dependent on the characteristics of that unit. If it has about the same productivity for both inputs it will move in a north-west/south-east direction across the substitution region but if it is e.g. very labour productive but less energy productive it will move in a south-west/north-east direction across the substitution region. In some cases the movement may be lightning-shaped. This can be seen in Figure A.2 where two examples of such strips are given by the shaded areas. Within each strip the units are partly utilised, while the utilisation rate for the unit in question obviously is zero to the left and one to the right of the strip, corresponding to the utilisation rates at the boundaries of each parallelogram.

The location of the activity regions (see Figure A.2) follows from a straightforward utilisation of the slope matrix. The substitution region may be filled up with activity regions by entering strips of parallelograms for each micro unit in turn. Choosing an arbitrary unit the units to be combined with this one are found by inspecting the corresponding column in the slope matrix.

The first unit to be combined with the chosen unit is the one with the largest absolute slope value, and then the other units are combined in descending order of the slope values. When a slope value is picked from the column, the corresponding parallelogram is formed by <u>subtracting</u> the full capacity input values of the unit in question from the previously obtained coordinate values in the substitution region, representing zero and full capacity utilisation, respectively, of the chosen unit. When a slope value is picked from the row the parallelogram is obtained by <u>adding</u> the full capacity input values. Thus, a partial utilisation strip changes direction each time the picking of consecutively decreasing slope values changes from row to column, or vice versa.

As an example, let us consider the shaded unit No. 9. The connecting line with the largest (absolute) slope, is with unit No. 1 found in the column of No. 9 in Table A.2.

The strip therefore starts in the direction of the origin. The next combination for unit No. 9 is Unit No. 6 found in the row of No. 9 and back to the column again, continuing with units No. 2, 3 and 11 in the direction towards the origin. Now all units with negative slopes have been combined with unit No. 9.

It is somewhat easier to follow the movement of unit No. 1. Since there are no negative signs in the column of this unit it means that it moves across the substitution region according to the ranking of absolute slope values in the row of this unit i.e. it starts with No. 9 and continues with No. 6, 8 and finally No. 5 at the lower boundary.

Isoquants

The isoquant we have chosen in this example goes through the activity regions of both the shaded units No. 1 and 9.

On the first 5 rows in Table A.3, the upper boundary of the sustitution region is built up and both the increments in labour, L, and energy, E, and the accumulated values are printed out. The last unit entered at the starting point of the isoquant (Unit No. 1) is utilised to 40.4 % of its capacity.

Starting at a chosen output level on the upper boundary, the last unit entered on the boundary is partially utilised. In Table A 3 this unit

Table A.3: The construction of an isoquant: 300 000 m^3 in 1980
L and E2 denote labour and total energy, respectively.

Line unit	Unit in	fraction before	fraction after	unit out	fraction before	fraction after	increments in L	increments in E	slope	coordinate values sum L	coordinate values sum E	comment
	4	ZERO	ONE	NONE			812.0	6053.0		812.0	6053.0	CONTOUR CORNER
	2	ZERO	ONE	NONE			370.0	3857.0		1182.0	9910.0	CONTOUR CORNER
	11	ZERO	ONE	NONE			268.0	2156.0		1450.0	12066.0	CONTOUR CORNER
	3	ZERO	ONE	NONE			503.0	3732.0		1953.0	15798.0	CONTOUR CORNER
	1	ZERO	0.404	NONE			250.6	2972.4		2203.6	18770.4	ISOQUANT START
1	9	ZERO	ONE	1	0.404	0.038	29.2	-1197.5	40.976	2232.8	17572.8	ISOQUANT CORNER
2	6	ZERO	0.100	1	0.038	ZERO	4.3	-154.8	35.633	2237.2	17418.0	ISOQUANT CORNER
3	6	0.100	ONE	9	ONE	0.050	11.5	-261.1	22.719	2248.7	17156.9	ISOQUANT CORNER
4	9	0.050	ONE	2	ONE	0.492	55.5	-539.7	9.723	2304.2	16617.2	ISOQUANT CORNER
5	8	ZERO	ONE	2	0.492	0.211	118.8	-211.9	1.783	2423.0	16405.4	ISOQUANT END

is No. 1. Referring to Table A.2, the algorithm inspects the figures in the column for the starting unit (No. 1), and the figures on the row for the same unit. For convenience, absolute values are used in this discussion. Thus the algorithm picks out the unit in the table yielding the steepest slope of the first isoquant segment by locating the largest figure either in the column or the row for the starting unit (No. 1). The slopes determining the isoquant are marked by rectangles in Table A.1.

The column for Unit No. 1 contains all utilised units, while the row consists of units which are not utilised.

If the largest figure is found in the column, the capacity utilisation of the starting unit is increased while the unit found in the column (unit numbers in front of the rows) is decreased.

If the largest figure is found in the row, the capacity utilisation of the starting unit is reduced, while the capacity utilisation of the unit found in the row (unit numbers at the top of the columns) is increased.

In the case of increased utilisation of the starting unit, the first isoquant corner point is reached when either the capacity of the starting unit is exhausted or the capacity utilisation of the decreasing unit reaches zero. When the capacity utilisation of the starting unit decreases, the corner point is reached when the utilisation of this unit reaches zero or the utilisation of the increasing unit reaches 100 %. At a corner, only one unit is partly utilised.

The first segment can, at most, be vertical because the boundary units are sorted according to increasing input coefficients of that input which is increasing along the isoquant towards the lower boundary. The actual length of the segment depends on the capacity of the activated units.

The next corner point is obtained by comparing the angles of all other units in the input coefficient space with the partly activated unit at the previously found corner point. The angle of the next line segment is then determined by the unit giving the steepest angle compared to the angle of the previous line segment, and so on, until the lower boundary is reached.

The successive angles, in the input coefficient space, between the units activated along the isoquant are the same as the slopes of the line segments in the input space.

The isoquant obtained according to the algorithm described above is piecewise linear, convex and it is as "close" to the origin as possible.

Let us now look at our example. The largest figure in Table A.2 for Unit No. 1 is 40.98, found in the row of this unit (the rectangle marked by 1 in Table A.2). Since it is found in a row, the capacity utilisation of this unit is reduced, from 40.4 % to 3.8 %. The largest figure, 40.98, is in the column of Unit No. 9. This means that the capacity utilisation of this unit is increased in this case from zero to 100.0 %, at the first corner point.

In our example the next line segment and corner point is found by inspecting the figures in the column of Unit No. 1 and the row for the same unit since Unit No. 1 is still the marginal one. The largest figure, not exceeding 40.98, is 35.63 in the row of Unit No. 1 and column of Unit No. 6. Since the largest figure is found in the row of Unit No. 1, the capacity utilisation of Unit No. 6 is increased from zero to 10 %. At the same time the capacity utilisation of Unit No. 1 is decreased from 3.8 % to zero. This yields the second corner point.

At the third step, the figures in the column and row of Unit No. 6 are inspected. It turns out that the largest figure, not exceeding 35.63 is 22.72, found in the column of Unit No. 6. This means that the capacity utilisation of Unit No. 6 now increases from 10 % to 100 %.

Since the figure 22.72 is found also in the row of Unit No. 9, the capacity utilisation of this unit decreases from 100 % to 5 %. This yields the third corner point.

Since Unit No. 9 is partially utilised at this corner point, the column and row of this unit are inspected again and the largest figure, not exceeding 22.72, is 9.72 in the column of Unit No. 9 and the row of Unit No. 2. Thus the capacity utilisation of Unit No. 9 increases again and this time up to 100 % at the fourth corner point, while that of Unit No. 2 decreases from 100 % to 49.2 %.

Since Unit No. 2 is now marginal we inspect the column and the row for this unit. We find that the largest figure next to 9.72 is 1.78 found in the row for Unit No. 2 and the column for Unit No. 8. This means that the capacity utilisation of Unit No. 2 is further reduced from 49.2 % to 21.1 % and the utilisation of Unit No. 8 increases from zero to 100 %. This yields the end point of the isoquant. This is seen by inspecting the column and the row for Unit No. 8. No negative number with an absolute value less than 1.78 can be found.

REFERENCES

AFRIAT, S.N. (1972): 'Efficiency Estimation of Production Functions', International Economic Review, Vol. 13, October 1972, 568-598.

AIGNER, D.J. and CHU, S.F. (1968): 'On Estimating the Industry Production Function', The American Economic Review, Vol. 58, No. 4, September 1968, 226-239.

AIGNER, D.J., AMEMIYA, T. and POIRIER, D.J. (1976): 'On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function', International Economic Review, 17, 377-396.

AIGNER, D.J., LOVELL, C.A.K. and SCHMIDT, P. (1977): 'Formulation and Estimation of Stochastic Production Function Models', Journal of Econometrics 6, 21-37.

AIRAKSINEN, T. (1974): 'Depreciation Rate and Capital Stock Measure: Empirical Experiments on a Plant Level in the Finnish Brewing Industry', Finnish Journal of Business Economics, Vol. 23.

AIRAKSINEN, T. (1977): A Dynamic Model of Interrelated Demand for Capital and Labour. A Theoretical Model with an Application to Time-series Cross-Section Data for Finnish Breweries in 1954-1972. Acta academiae oeconomicae Helsingensis, Series A:20, Helsinki 1977.

BALLANCE, R. and SINCLAIR, S. (1983): Collapse and Survival: Industry Strategies in a Changing World, World Industry Studies 1. George Allen & Unwin, London, 1983.

BAUER, P.W. (1984): An Analysis of Multi-Product Firm Technology and Efficiency Using Panel Data, manuscript. Department of Economics, University of North Carolina.

BINSWANGER, H. (1974a): 'A Cost Function Approach to the Measurement of Elasticities of Substitution', American Journal of Agricultural Economics, Vol. 56.

BINSWANGER, H. (1974b): 'The Measurement of Technological Change Biases with Many Factors of Production', American Economic Review, Vol. 64, 964-976.

BLACKORBY, C. - LOVELL, K. - THURSBY, M. (1976): 'Extended Hicks Neutral Technical Change', The Economic Journal, Vol. 86, 845-851.

BOSTON CONSULTING GROUP (1974): The Experience Curve - Reviewed, Boston.

BOSWORTH, D.L. (1976): Production Functions. A Theoretical and Empirical Study. Saxon House/Lexington Books, Glasgow 1976.

BROEK, J. - FØRSUND, F. - HJALMARSSON, L. - MEEUSEN, W. (1980): 'On the Estimation of Deterministic and Stochastic Frontier Production Functions: A Comparison', Journal of Econometrics 13, 117-138.

BROUWER, M. (1981): 'The European Beer Industry: Concentration and Competiton', in de JONG, H. (ed.) The Structure of European Industry, The Hague/Boston/London, 39-56.

CARLSSON, B. (1972): 'The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries 1968', The Swedish Journal of Economics, Vol. 74, No. 4, December 1972, 468-485.

COWING, T., REIFSCHNEIDER, D. and STEVENSON, R. (1983): 'A Comparison of Alternative Frontier Cost Function Specifications', Ch. 4 in Developments in Econometric Analyses of Productivity. DOGRAMACI, A. (ed.). Kluwer-Nijhoff.

DIEWERT, W.E. (1971): 'An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function', Journal of Political Economy, Vol. 79.

DIEWERT, W.E. (1974): 'Applications of Duality Theory' in Frontiers of Quantitative Economics, Volume II, edited by IINTRILIGATOR, M.D. and KENDRICK, D.A., Amsterdam: North Holland Publishing Company.

DIEWERT, W.E. (1976): 'Exact and Superlative Index Numbers', Journal of Econometrics.

DIEWERT, W.E. (1980): 'Aggregation Problem in the Measurement of Capital' in Usher, D. (ed.) (1980): The Measurement of Capital, The University of Chicago Press, Chicago.

DIEWERT, W.E. (1981a): 'The Theory of Factor Productivity Measurement in Regulated Industries', in COWING and STEPHENSON, eds.: Productivity Measurement in Regulated Industries, Academic Press, New York.

DIEWERT. W.E. (1981b): 'The Economic Theory of Index Numbers: A Survey' in DEATON, A. (ed.): Essays in the Theory and Measurement of Consumer Behaviour in honour of Sir Richard Stone. Cambridge University Press, Cambridge.

DIEWERT, W.E. (1982): 'Duality Approaches to Microeconomic Theory', in ARROW, K.J. and INTRILIGATOR, M.D., eds.: Handbook of Mathematical Economics, North Holland Publishing Company, Amsterdam, pp. 535-591.

EIDE, E. (1979): Engineering Production and Cost Functions for Tankers. Amsterdam-Oxford-New York: Elsevier, 1979.

ELLIOT, J.E. (1980): 'Marx and Schumpeter on Capitalism's Creative Destruction: A Comparative Restatement', Quarterly Journal of Economics, 95, 45-68.

ETLA (1984): Kansantalouden kehitysnäkymät 1984-1988. Medium Term Economic Prospects, with English summary. Helsinki.

FARRELL, M.J. (1957): 'The Measurement of Productive Efficiency', Journal of the Royal Statistical Society, Series A (General), 120, Part 3, 1957, 253-290.

FARRELL, M.J. and FIELDHOUSE, M. (1962): 'Estimating Efficient Production Functions under Increasing Returns to Scale', Journal of the Royal Statistical Society, Series A (General), 125, Part 2, 1962, 252-267.

FISCHER, K.H. (1984): Die Messung von totaler Faktorproduktivität, Effizienz und technischem Fortschritt. Bonner Betriebswirtschaftliche Schriften 16, Bonn.

FISHER, F.M. (1968a): 'Embodied Technology and the Aggregation of Fixed and Movable Capital Goods', Review of Economic Studies, Vol. 35, 429-442.

FISHER, F.M. (1968b): 'Embodied Technology and the Existence of Labour and Output Aggregates', Review of Economic Studies, Vol. 35, 429-442.

FISHER, F.M. (1969): 'The Existence of Aggregate Production Functions', Econometrica, Vol. 37, 553-577.

FRISCH, R. (1965): Theory of Production. D. Reidel, Dordrecht, Holland.

FUSS, M.A. (1977): 'The Structure of Technology over Time: A Model for Testing the "Putty-clay" Hypothesis', Econometrica, Vol. 45, No. 8, November 1977, 1797-1821.

FUSS, M. - McFADDEN, D. (eds.) (1978): Production Economics: A Dual Approach to Theory and Applications, North-Holland Publishing Company, Amsterdam.

FÄRE, R., GROSSKPF, S. and LOVELL, C.A.K. (1983): 'The Structure of Technical Efficiency', Scandinavian Journal of Economics, 85(2), 181-190.

FÄRE, R. and LOVELL, K. (1978): 'Measuring the Technical Efficiency of Production', Journal of Economic Theory, 19, 150-162.

FØRSUND, F. (1974): Studies in the Neo-classical Theory of Production. Memorandum from Institute of Economics, University of Oslo, Oslo.

FØRSUND, F.R. (1985): 'Comment' (on P. SCHMIDT's article 'Frontier Production Functions'), Econometric Reviews, 4.

FØRSUND, F.R., GAUNITZ, S., HJALMARSSON, L. and WIBE, S. (1980): 'Technical Progress and Structural Change in the Swedish Pulp Industry 1920-74', in The Economics of Technological Progress, ed. by T. Puu and S. Wibe, London: MacMillan.

FØRSUND, F.R. and HJALMARSSON, L. (1974): 'On the Measurement of Productive Efficiency', Swedish Journal of Economics, 76, 141-154.

FØRSUND, F.R. and HJALMARSSON, L. (1979): 'Frontier Production Function and Technical Progress: A Study of General Milk Processing in Swedish Dairy Plants', Econometrica, 47, 883-900.

FØRSUND, F.R. and HJALMARSSON, L. (1983): 'Technological Progress and Structural Change in the Swedish Cement Industry 1955-1979', Econometrica.

FØRSUND, F.R. and HJALMARSSON, L. (1984): Analysis of Industrial Structure: A Production Function Approach, IUI Working Papers No. 135.

FØRSUND, F.R. and JANSEN, E. (1974): Average Practice and Best Practice Production Functions with Variable Scale Elasticity: An Empirical Analysis of the Structure of the Norwegian Mechanical Pulp Industry, Memorandum from the Institute of Economics, University of Oslo, Oslo 1974.

FØRSUND, F.R. and KNOX LOVELL, C.A. and SCHMIDT, P. (1980): 'A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement', Journal of Econometrics, No. 1.

GABRIELSSON, A. (1970): Koncentration och skalekonomi inom malt- och läskedrycksindustrin under 1950- och 1960-talet, Forskningsrapporter från Kulturgeografiska institutionen, Nummer 20, Uppsala Universitet.

GOLLOP, M. - JORGENSON, D. (1980): 'U.S. Productivity Growth by Industry, 1947-73' in KENDRICK, J. - VACCARA, B. (eds.): New Developments in Productivity Measurement and Analysis, National Bureau of Economic Research, The University of Chicago Press, Chicago and London.

GORMAN, W. (1968): 'Measuring the Quantities of Fixed Factors' in Wolfe, J. (ed.): Value, Capital and Growth: Papers in Honour of Sir John Hicks, Aldine.

GREENE; W. (1980): 'On the Estimation of Flexible Frontier Production Model', Journal of Econometrics 13, 101-115.

GREENE, W.H. (1982): 'Maximum Likelihood Estimation of Stochastic Frontier Production Models', Journal of Econometrics, 18, 285-289.

GRILICHES, Z. and RINGSTAD, V. (1971): Economies of Scale and the Form of the Production Function, North-Holland Publishing Company, Amsterdam 1971.

GROSSE, A. (1953): 'The Technological Structure of the Cotton Textile Industry', in Studies in the Structure of the American Economy, (ed. W. Leontief).

GUILKEY, D.K. and KNOX LOVELL, C.A. (1980): 'On the Flexibility of the Translog Approximation', International Economic Review, Vol. 21, No. 1 (Feb.), pp. 51-67.

HATTEN, K. - SCHENDEL, D. (1976): Heterogeneity within an Industry: Firm Conduct in the U.S. Brewing Industry 1952-1971. Purdue University, Paper No. 581.

HATTEN, K. - SCHENDEL, D. - COOPER, A. (1976): A Strategic Model of the U.S. Brewing Industry 1952-1971. Purdue University, Paper No. 580.

HECKSCHER, E.F. (1918): Svenska produktionsproblem, p. 13. Bonniers, Stockholm.

HELPER, S. - LAZONIK, W. (1984): Learning Curves and Productivity Growth in Institutional Context. Harvard Institute of Economic Research, Harvard University, Cambridge, Massachusetts, Discussion paper No. 1066.

HILDENBRAND, W. (1981): 'Short-run Production Functions Based on Microdata', Econometrica, Vol. 49, 1095-1124.

HILDENBRAND, K. (1982): 'Numerical Computation of Short-Run Production Functions', in Quantitative Studies on Production and Prices, ed. by W. Eichhorn, R. Henn, K. Neumann and R.W. Shephard. Physica-Verlag, Vienna.

HJALMARSSON, L. (1973): 'Optimal Structural Change and Related Concepts', Swedish Journal of Economics, 75 (2), 176-192.

HJALMARSSON, L. (1974): 'The Size Distribution of Establishments and Firms Derived from an Optimal Process of Capacity Expansion', European Economic Review, 5, 123-140.

HJALMARSSON, L. (1975): Studies in a Dynamic Theory of Production and its Applications, Department of Economics, University of Gothenburg, Memorandum No. 50.

HOUTHAKKER, H.S. (1955): 'The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis', Review of Economic Studies, Vol. 23, 27-31.

HUAN, C.J. (1984): 'Estimation of Stochastic Frontier Production Function and Technical Inefficiency via the EM Algorithm', Southern Economic Journal, 50, 847-856.

INTRILIGATOR, M.D. (1971): Mathematical Optimization and Economic Theory. Englewood Cliffs, N.J., Prentice-Hall.

JOHANSEN, L. (1972): Production Functions - An Integration of Micro and Macro, Short Run and Long Run Aspects. North Holland Publishing Company, Amsterdam.

JONDROW, J., KNOX LOVELL, C.A., MATEROW, I.S., SCHMIDT, P. (1982): 'On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model', Journal of Econometrics, Vol. 19, 233-238. North-Holland Publishing Company.

JORGENSON, D.W. (1967): 'The Theory of Investment Behavior' in FERBER, R. (ed.): Determinants of Investment Behavior. Columbia University Press.

JORGENSON, D.W. (1974): 'The Economic Theory of Replacement and Depreciation', in SELLEHAERTS, W. (ed.): Econometrics and Economic Theory: Essays in Honor of Jan Tinbergen. International Arts and Sciences Press, White Plains, New York.

JORGENSON, D. - GRILICHES, Z. (1967): 'The Explanation of Productivity Change', Review of Economic Studies.

KARKO, J. (1986): On the Measurement of Technical Differences - A Cost Theoretical Approach. ETLA, Discussion Papers No. 207, 1986.

KLEIN, L.R. (ed.) (1971): Essays in Industrial Economics, Vol. III, Economics Research Unit, Department of Economics, Wharton School of Finance and Commerce, University of Pennsylvania, Studies in Quantitative Economics No. 5, Pennsylvania 1971.

KOPP, R. (1981): 'The Measuremet of Productive Efficiency: A Reconsideration, The Quarterly Journal of Economics, 477-503.

KOPP, R. - SMITH, V. (1980): 'Frontier Production Function Estimates for Steam Electric Generation: A Comparative Analysis', Southern Economic Journal, 1049-1059.

KRELLE, W. (1969): Produktionsheorie. Tübingen

KUMBHAKAR, S.C. (1985a): Specification of Input-specific Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers. University of Southern California, Modelling Research Group Working Paper #M8512.

KUMBHAKAR, S.C. (1985b): Econometric Models for Input-Specific Technical and Allocative Inefficiency: U.S. Class 1 Railroads, University of Southern California, Modelling Research Group, Working Paper #M8521.

LANGE, O. - TAYLOR, F. (1938): On the Economic Theory of Socialism, (ed.) B. Lippincott. University of Minnesota Press, Minneapolis.

LASTIKKA, P. (1977): Tutkimus Suomen teollisuustyöntekijöiden työvoimapiireittäisistä palkkaeroista vuosina 1960-1971. ETLA Sarja C 12, Helsinki.

LAU, L.J. (1974): 'Applications of Duality Theory: Comments to Diewert' in INTRILIGATOR, M.D. and KENDRICK, D.A. (eds.): Frontiers of Quantitative Economics, Vol. II. North-Holland Publishing Company, Amsterdam.

LAU, L.J. (1978): 'Applications of Profit Functions', in FUSS, M. and McFADDEN, D. (eds.): Production Economics: A Dual Approach to Theory and Applications, Vol. 1. North-Holland Publishing Company, Amsterdam.

LAU, L.J. and YOTOPOULOS, P.A. (1971): 'A Test for Relative Efficiency and Application to Indian Agriculture', American Economic Review, Vol. 61, No. 1, March 1971.

LEE, L.-F. (1983): 'A Test for Distributional Assumptions for the Stochastic Frontier Functions', Journal of Econometrics, 22, 245-257.

LEHTONEN, P. (1976): On the Estimation of Statistical Cost Functions: An Empirical Study of Production and Cost Functions in Finnish restaurants, Acta academiae oeconomicae Helsingensis, Series A:19, Helsinki 1976.

LEIBENSTEIN: H. (1966): 'Allocative Efficiency vs. X-Efficiency', The American Economic Review, Vol. 56, June 1966.

LEIBENSTEIN, H. (1975): 'Aspects of the X-Efficiency Theory of the Firm', *The Bell Journal of Economics*, Vol. 6, Autumn 1975.

LERNER, A. (1944): *The Economics of Control*. Macmillan, New York.

MALTZAN, B. (1978): 'Average'-Produktionsfunktionen und Effizienzmessung über 'frontier production functions'. *Bonner Betriebswirtschaftliche Schriften 3*, Bonn 1978.

MANSKI, C. (1984): 'Adaptive Estimation of Non-Linear Regression Models', *Econometric Reviews*, 3, 145-194.

MARCHAND, M., PESTIEAU, P. and TULKENS, H. (eds.) (1984): *The Performance of Public Enterprises - Concepts and Measurement*. Studies in Mathematical and Managerial Economics, Vol. 33. North Holland.

MARSCHAK, J. and ANDREWS, W.M. Jr. (1944): 'Random Simultaneous Equations and the Theory of Production', *Econometrica*, Vol. 12, 143-205.

McFADDEN, D. (1978): 'Cost, Revenue and Profit Functions', in FUSS, M. and McFADDEN, D. (eds.): *Production Economics: A Dual Approach to Theory and Applications*, Vol. 1. North-Holland Publishing Company, Amsterdam.

McKENZIE, E. (1937): *Planned Society: Yesterday Today Tomorrow*. Prentice Hall, New York.

MEEUSEN, W. - van den BROECK, J. (1977a): 'Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error', *International Economic Review*, 18, 435-444.

MEEUSEN, W. - van den BROECK, J. (1977b): 'Technical Efficiency and Dimension of the Firm: Some Results on the Use of Frontier Production Functions', *Empirical Economics*, 2, 109-122.

MELFI, C.A. (1984): *Technical and Allocative Efficiency in Electric Utilities*, Working Paper, Department of Economics, Indiana University.

MOSS, S. (1984): 'The History of the Theory of the Firm from Marshall to Robinson and Chamberlin: The Source of Positivism in Economics', *Economica*, 51, 307-318.

MUELLER, J. - SCHWALBACH, J. (1977): *Structural Change in West Germany's Brewing Industry: Some Efficiency Considerations*. International Institute of Management, Discussion Papers 77-115.

MUYSKEN, J. (1979): *Aggregation of Putty-Clay Production Functions*. A methodological study of the distribution approach, applied to the Japanese cotton spinning industry. Groningen 1979.

NADIRI, M. (1970): 'Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey', *Journal of Ecnomic Literature 8 (4)*, 1137-1177.

PANIC, M. (ed.) (1976): *The UK and West German Manufacturing Industry 1954-72*, NEDO Monograph 5. London.

PHELPS, E. (1963): 'Substitution, Fixed Proportions, Growth and Distribution', International Economic Review, 4, pp. 265-288.

PITT, M. and LEE, LUNG-FEI (1981): 'The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry', Journal of Development Economics, 9, 43-64.

PORTER, M.E. (1979): 'The Structure within Industries and Companies' Performance', Review of Economics and Statistics, 61, 214-227.

PORTER, M.E. (1980): Competitive Strategy: Techniques for Analysing Industries and Competition. Macmillan, New York.

PRATTEN, C.F. (1971): Economies of Scale in Manufacturing Industry. Cambridge University Press.

PRATTEN, C.F. (1976): A Comparison of the Performance of Swedish and UK Companies, University of Cambridge, Department of Applied Economics, Occasional Papers 47, Cambridge University Press, Cambridge.

PRESSMAR, D.N. (1971): Kosten- und Leistungsanalyse im Industriebetrieb. Wiesbaden.

PUPUTTI, R. (1983): Teollisten yritysten toimialalle tulo ja siltä poistuminen. Salterin teorian sovellutus Suomen panimoteollisuudessa. Master's Thesis in Economics, Helsinki School of Economics.

RIBRANT, G. (1970): Stordriftsfördelar inom industriproduktionen (Economies of Scale in Industrial Production), SOU (The Swedish Government Committee on the Concentration of Economic Power) 1970:3, Stockholm 1970.

RICHMOND, J. (1974): 'Estimating the Efficiency of Production', International Economic Review, Vol. 15, No. 2, June 1974, 515-521.

ROBBINS, L. (1928): 'The Representative Firm', Economic Journal, 36, 387-404.

RODAS, S. - HUMBERG, R. (1980): Produktivitet i verkstadsföretag: En jämförelse Finland-Sverige. SITRA, Serie B No. 54.

RUSSELL, R.R. (1985): 'Measures of Technical Efficiency', Journal of Economic Theory, 35, 109-126.

SALTER, W.E.G. (1960): Productivity and Technical Change. Cambridge University Press, Cambridge 1960.

SAMUELSSON, P. (1947): Foundations of Economic Analysis. Harvard University Press, Cambridge, Mass.

SAMUELSSON, P. - SWAMY, P. (1974): 'Invariant Economic Index Numbers and Canonical Duality', American Economic Review, 566-593.

SATO, K. (1975): Production Functions and Aggregation. Amsterdam.

SATO, R. (1981): Theory of Technical Change and Economic Invariancy by Application of Lie Groups. Academic Press, San Francisco.

SCHMIDT, P. (1976): 'On the Statistical Estimation of Parametric Frontier Production Functions', Review of Economics and Statistics, 58, 238-239.

SCHMIDT, P. (1985): 'Frontier Production Functions', Econometric Reviews 4.

SCHMIDT, P. - LOVELL, K. (1979): 'Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers', Journal of Econometrics, 9, 343-366.

SCHMIDT, P. and SICKLES, R.C. (1984): 'Production Frontiers and Panel Data', Journal of Business and Economic Statistics, 2, 367-374.

SCHWALBACH, J. (1981): Struktur und Wettbewerb in der deutschen Brauereiindustrie, International Institute of Management, Discussion Papers 81-25.

SCHWALBACH, J. (1984 a): Ausmass und Entwicklung von Grössenvorteilen in der deutschen Bier- und Zementindustrie, International Institute of Management, Discussion Papers 84-13.

SCHWALBACH, J. (1984 b): Marktstruktur und strategisches Verhalten der deutschen Bierindustrie, International Institute of Management, Discussion Papers 84-12.

SCHWALBACH, J. (1984 c): Rentabilitätsentwicklung deutscher Brauereiaktiengesellschaften, International Institute of Management, Discussion Papers 84-22.

SEERINGER, W. (1975 a): Ein Modell zur optimalen Betriebsgrössen und Standortplanung am Beispiel der Brauindustrie in der Bundesrepublik Deutschland, International Institute of Management, Discussion Papers 75-1.

SEERINGER, W. (1975 b): Zur Frage der optimalen Betriebsgrösse von Braustätten, International Institute of Management, Discussion Papers 75-2.

SEIERSTAD, A. (1985): 'Properties of Production and Profit Functions Arising from the Aggregation of a Capacity Distribution of Micro Units', in FØRSUND, F., HOEL, M. and LONGVA (Eds.): Production, Multi-Sectoral Growth and Planning. Elsevier Science Publishers B.V. (North-Holland), 1985.

SEITZ, W.D. (1970): 'The Measurement of Efficiency Relative to a Frontier Production Function', American Journal of Agricultural Economics, Vol. 52, 505-511.

SEITZ, W.D. (1971): 'Productive Efficiency in the Steam-Electric Generating Industry', Journal of Political Economy, Vol. 79, 878-886.

SHEN, T.Y. (1984): 'The Estimation of X-Inefficiency in Eighteen Countries', Review of Economics and Statistics, Feb. 1984.

SHEPHARD, R. (1953): Cost and Production Functions. Princeton, N.J.: Princeton University Press.

SHEPHARD, R.W. (1963): Cost and Production Functions. Princeton: Princeton University Press.

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

SIMULA, M. (1983): Productivity Differentials in the Finnish Forest Industries. Suomen Metsätieteellinen Seura, Helsinki 1983.

STERNER; T. (1985): Energy Use in Mexican Industry. Ekonomiska studier utgivna av nationalekonomiska institutionen vid Göteborgs universitet, 15. Göteborg.

STEVENSON, R.E. (1980): 'Likelihood Functions for Generalized Stochastic Frontier Estimation', Journal of Econometrics, 13, 57-66.

SUMMA, T. (1971): Tutkimus Suomen panimo- ja virvoitusjuomateollisuuden kansainvälisestä kilpailukyvystä. SITRA A No. 8.

SUMMA, T. (1972): Panimoteollisuuden kehitys vuosina 1959-1970, Alkoholipolitiikka 1972:1.

SUMMA, T. (1976): On Describing and Projecting Industrial Development, ETLA, Discussion Papers No. 9.

SUMMA, T. (1979): Teollisen toimialan tehokkuusrakenne ja sen muutokset. Salterin teoria ja sen soveltuvuus toimialan kehitysprosessin tarkasteluun, ETLA C 16.

SUMMA, T. (1985): Development of Intra-Industry Efficiency. ETLA, C 36.

SUMMA, T., FØRSUND, F.R., HJALMARSSON, L., KARKO, J. and EITRHEIM, Ø. (1985): An Intercountry Comparison of Productivity and Technical Change in the Nordic Cement Industry. ETLA, B 44.

SWAMY, P.A. (1971): Statistical Inference in Random Coefficient Regression Models. Berlin, Springer Verlag.

SVENNILSON, I. (1944): 'Industriarbetarnas växande avkastning i belysning av svenska erfarenheter', särtryck ur Studier i ekonomi och historia, tillägnade Eli F. Heckscher, 24.11.1944.

SÄRSKILDA NÄRINGSPOLITISKA DELEGATIONEN (1979): Vägar till ökad välfärd. Expertbilaga 1 till Ds Ju 1979:1, Ds Ju 1979:2.

TEAGUE, J. - EILON, S. (1973): 'Productivity Measurement: A Brief Survey', Applied Economics, 5, 133-145.

THURSBY, M. and KNOX LOVELL, C.A. (1978): On the Flexibility of the Translog Approximation, Discussion Paper No. 78-07, Department of Economics, University of North Carolina at Chapell Hill.

TIMMER, C.P. (1970): On Measuring Technical Efficiency. Food Research Institute Studies in Agricultural Economics, Trade and Development, Vol. 9, No. 2, 1970.

TIMMER, C.P. (1971): 'Using a Probabilistic Frontier Production Function to Measure Technical Efficiency', Journal of Political Economy, Vol. 79, No. 4, July/August 1971, 776-794.

TODA, Y. (1976): 'Estimation of a Cost Function When the Cost Is Not Minimum: The Case of Soviet Manufacturing Industries, 1958-71', Review of Economics and Statistics, 58, No. 3, 259-268.

TODA, Y. (1977): 'Substitutability and Price Distortion in the Demand for Factors of Production: An Empirical Estimation', Applied Economics, 9, No. 2, 203-217.

TODD, D. (1971): The Relative Efficiency of Small and Large Firms. Research Report No. 18, Committee of Inquiry on Small Firms, London: H.M.S.O.

TODD, D. (1985): 'Productive Performance in West German Manufacturing Industry 1970-80: A Farrell Frontier Characterisation', The Journal of Industrial Economics, Volume 33, March 1985, 295-316.

UEBE, G. (1976): Produktionstheorie. Lecture Notes in Economcis and Mathematical Systems, No. 114. Springer, Heidelberg.

VARIAN, H. (1978): Microeconomic Analysis. Norton & Company, New York.

VARIAN, H. (1984): 'The Nonparametric Approach to Production Analysis', Econometrica, 52, 579-597.

VARTIA, Y.O. (1973): Estimation of the Parameters of the Beta Distribution by Maximum Likelihood Method, ETLA A 1.

VARTIA, Y.O. (1976): Relative Changes and Index Numbers. ETLA, Series A 4, Helsinki.

VINELL, L. (1981): The Hagfors Strip Mill: A Study of Productivity Change in a Hot Rolling Strip Mill.

VUORI, S. (1981): Yritysten toimialoittaiset kokojakautumat Suomessa. English summary: Size Distributions of Finnish Enterprises by Branches. ETLA, Series B 30, Helsinki.

WALDMAN, D.M. (1982): 'A Stationary Point for the Stochastic Frontier Likelihood',Journal of Econometrics, 18, 275-279.

WARD, M. (1976): The Measurement of Capital. The Methodology of Capital Stock Estimates in OECD Countries. OECD, Paris.

WIBE, S. (1980): Teknik och aggregering i produktionsteorin. Svensk järnhantering 1850-1975; en branschanalys. University of Umeå, Umeå Economics Studies Nr 63, Umeå.

WOHLIN, L. (1970): Skogsindustrins strukturomvandling och expansionsmöjligheter (Forest-Based Industries: Structural Change and Growth Potentials), Industriens Utredningsinstitut, Stockholm.

YLÄ-LIEDENPOHJA, J. - TÖRMÄ, H. (1984) The Financial Cost of Capital in Finnish Manufacturing 1960-83. Unpublished manuscript, University of Jyväskylä, Department of Economics and Management.

ZELLNER, A. - REVANKAR, N. (1969): 'Generalized Production Functions', Review of Economic Studies, Vol. 36, 241-250.

ÅSELIUS, H. (1957): Investments and Productivity in the Iron and Steel Industry. Index, June, Supplement, 3-13.

ÖSTERBERG, E. (1974): Alkon panimopolitiikka vuosina 1948-1972. Helsinki: Alkoholipoliittisen tutkimuslaitoksen tutkimusseloste no. 79.