IUI CONFERENCE REPORTS 1978:2

BO CARLSSON,
GUNNAR ELIASSON, ISHAQ NADIRI, EDITORS.

The Importance of Technology and the Permanence of Structure in Industrial Growth

Proceedings of a Symposium at IUI, Stockholm, July 18–19, 1977

IUI

THE INDUSTRIAL INSTITUTE FOR
ECONOMIC AND SOCIAL RESEARCH, STOCKHOLM.

The Industrial Institute for Economic and Social Research

is an independent non-profit research institution, founded in 1939 by the Swedish Employers' Confederation and the Federation of Swedish Industries.

Objectives

To carry out research into economic and social conditions of importance for industrial development in Sweden.

Activities

The greater part of the Institute's work is devoted to long-term problems, especially to long-term changes in the structure of the Swedish economy particularly within manufacturing industry. This also includes continuous studies of the development of private consumption in Sweden and projections of demand for various products. Research results are published in the series issued by the Institute.
Along with the long-term research work the Institute carries out investigations concerning special problems and performs certain services to industrial enterprises, organizations, governmental agencies, etc.

Board

Address

Industriens Utredningsinstitut
Grevgatan 34, 5 tr, S-114 53 Stockholm, Sweden
Tel. 08-63 50 20

ISBN 91-7204-084-X

The Importance of Technology and the Permanence of Structure in Industrial Growth

IUI The Industrial Institute for Economic and Social Research

IUI Conference Reports
1978:2

THE IMPORTANCE OF TECHNOLOGY AND THE PERMANENCE OF STRUCTURE IN INDUSTRIAL GROWTH

Proceedings of a Symposium
at IUI, Stockholm, July 18–19, 1977

Editors:
Bo Carlsson, Gunnar Eliasson and Ishaq Nadiri

Distributor:
Almqvist & Wiksell International,
Stockholm

ISBN 91–7204–084–X
GOTAB Stockholm 1978 59663

CONTENTS

Foreword

Introduction

A Vintage Model of Swedish Economic Growth from 1870 to 1975 1
Ragnar Bentzel, IUI, Stockholm

A Dynamic Model of Research and Development Expenditure 5
Ishaq Nadiri, New York University

Relative Price Change and Industrial Structure – The "Norwegian Case" 7
Gunnar Eliasson, IUI, Stockholm

Choice of Technology in the Cement Industry – A Comparison of the United States and Sweden 9
Bo Carlsson, IUI, Stockholm

Technical Change in the Swedish Hydro Power Sector 1900–1975 14
Anders Grufman, IUI, Stockholm

Technical Progress and Structural Efficiency of Swedish Dairy Plants 16
Finn R. Førsund, Oslo University, and Lennart Hjalmarsson, IUI, Stockholm

Production Frontiers of Individual Firms in Swedish Manufacturing 1975 and 1976 18
Jim Albrecht, Columbia University, New York

A Dynamic Formulation of the Law of Diminishing Returns 20
Rolf Färe, University of California, Berkeley

On Ray-Homothetic Production Functions 22
Rolf Färe, University of California, Berkeley,
Leif Jansson, IUI, Stockholm, and C.A. Knox Lovell,
University of North Carolina at Chapel Hill

FOREWORD

Industrial structure and technical change have long been a traditional field of research at the Institute. Professor Nadiri's visit to the Institute during the summer of 1977 afforded us an opportunity to arrange this small international seminar and to pool ongoing IUI research around this theme. The seminar also offered an excellent way of reviewing the results obtained so far and to contrast them with outside research. We found this particularly important in order to get an early and firm grasp of what we know about the importance of technical change in the Swedish growth process for the current joint research venture on the technical competence of Swedish industry with the Royal Swedish Academy of Engineering Sciences. The sample of research activities reported on at the seminar from IUI staff and outside researchers combines theoretical and empirical analysis. The coverage of methods applied to a common theme was very broad, and the experience gained has been useful for the direction of efforts in the ongoing joint technology project.

The Institute wants to thank all outside participants for their contributions, and in particular Professor Nadiri, New York University, who is a coeditor of this conference volume.

Stockholm, August 1978

Gunnar Eliasson

INTRODUCTION

Attempts to explain the growth process using aggregate models and data have turned out to be somewhat unrewarding. Whichever way received theory has been molded, growth has been explained as either depending directly upon time or upon some exogenous coefficient that sets the pace of a central growth factor like production capacity or capital accumulation. It has become increasingly evident that our specification of production relationships needs improvements to escape such confining framework. Such improvements are needed in all areas: in theoretical framework, in statistical measurements, and in data and estimation. In this volume an effort is made to contribute in each of these fields. However, emphasis is put on integrating engineering information with economic reasoning in the context of specific industry studies to illustrate their evolution. For what is important is *not* further refinement of those structures that we already have, but a major infusion into economics of technical and engineering *knowledge*. In particular we need to specify what constitutes productivity change at the plant level more exactly and to do so with a sufficient degree of systematization to make generalizations possible.

The theme of "how to measure and analyze..." is common to all the papers, but the approaches actually used are very different. We think that this very wide range of methods, from a careful down-to-earth investigation of what is really going on in one particular industry (Carlsson) to the upper ranges of theory (Färe) makes extremely good sense in stressing the various aspects of the same problem.

The volume includes four sets of papers: the first set (Bentzel and Eliasson) addresses more general problems of change in the total economy in response to dramatic shifts in relative prices and to sectoral technical changes. The second set of papers (Färe–Jansson–Lovell, Albrecht and Nadiri) is methoddological in nature, suggesting new ways of measuring technical change and the underlying production process. The third set of papers (Carlsson, Grufman, Førsund–Hjalmarsson) relates to specific industry applications in which the economic principles and engineering information and specifications are explicitly taken into account. The penultimate paper (Färe) is a theoretical description of a dynamic formulation of the law of diminishing returns. The insights provided in this paper could contribute greatly to the estimation and explanation of the growth process both at the aggregate and industry level.

The main findings of these sets of papers can be stated briefly.

Bentzel sets out to capture over 100 years of Swedish economic growth (1870–1975) in a simple, one-sector, vintage production function model for the entire economy of Sweden, excluding the public sector, with a disembodied technical labour-augmenting factor. He reaches the conclusion that there was a fairly significant structural change in the Swedish economy around 1930; the length of life of equipment has declined dramatically in recent years suggesting a strong increase in equipment turnover. A secular decline in the output/capital ratio of new vintages is evident which commenced in the 1930's and has been accelerating since the middle of the 1960's. The social marginal productivity of capital which stayed around 20 % between 1870 and 1930 declined substantially to 12 % in 1975. Finally, the variability of the Swedish economy's growth rate during the 100-year period has been directly related to growth of net investment and replacement of older vintage equipment. On the whole Bentzel's results set the stage for a fairly pessimistic scenario of the future of Sweden if we cannot generate a sudden jump either in the investment ratio or in the labor-augmenting technical factor or (preferably) in both.

The influence of a sudden change in relative prices on industrial structure is investigated by *Eliasson* in an individual firm based macro simulation model of the Swedish economy, developed at the Institute. The experimental setting mimics the Norwegian experience of a sudden discovery of a new "land rent" (North Sea Oil) followed by a dramatic price hike in the same sector. This possibility, when applied in somewhat extreme form to the entire Swedish raw material producing sector, appears as a mixed blessing to the economy at large. Even for such a "mature" industrialized economy as Sweden, the excessive wage inflation occasioned by the "discovery" and the subsequent foreign price increase proves strongly detrimental to rates of return and growth in other sectors.

The explanation lies in the disturbances in the market price signalling and allocation functions of the markets, caused by the size and suddenness of the price change. It is interesting that Bentzel arrived at essentially the same results by looking at the production side only. This analysis reinforces the gloomy outlook for the Swedish economy and focuses attention on at least one originating factor, namely inflation.

In the second set of papers two methodological estimation issues are addressed: one is whether technological innovations are endogenous and whether they affect other inputs such as labor and capital; the other issue refers to the returns to scale of the production process.

The paper by *Nadiri* explores the interrelationships between a firm's employment, capital accumulation and research and development decisions.

He uses a production function adjusted by the factor utilization rate with R & D stocks explicitly included as an input and a disembodied technical shift factor. Derived demand functions are formulated that obviously suggest that inputs respond to changes in output and that there are strong feedbacks among input decisions and relative factor prices. Individual firm data (62 firms) for the years 1965–72 from the NBER Compustat tapes have been used in a combined cross section and time series analysis. The results turn out to be size.

Nadiri finds that changes in output and prices have had a strong influence on the chosen combination of labor, capital and R & D inputs and that the decision on how to mix factors is complex. In particular he observes a complementarity between capital goods stocks and R & D expenditures and a substitutability between capital goods and employment.

The ray-homothetic production function analyzed and estimated for the first time on data for U.S. transportation industries by state allows *Färe-Jansson-Lovell* to link returns to scale directly with both output (the homothetic side) and the factor input mix (the ray-homogeneous side). They find strong support for the combined homothetic and ray-homogeneous formulation. A large part of U.S. transport production is found to take place in an interval where increasing returns to scale obtain. Hence average actual output is substantially below what is technically optimal.

The *Albrecht* paper introduces a new data base on capacity utilization that makes it possible to estimate production frontiers and describe the structure of the production system on the format used in the micro-to-macro model of the Eliasson paper. Albrecht explains the estimation technique and applies it to data for the years 1975 and 1976 on more than 200 Swedish production units. The idea of the estimation technique in Albrecht's paper is to exploit data on the presence of labor redundancy in industry. This is often substantial, and suggests implications quite different from those to be expected were firms always operating on their frontiers.

The third set of papers investigates the response patterns of specific industries to changes in relative prices, causes of structural changes in some industries, and factors behind technical progress in certain industries. The papers by *Grufman* and *Førsund-Hjalmarsson* suggest that in both hydro-electricity and milk production, the efficiency of best practice plants increase faster over time than the corresponding average for the industry. From a growth point of view this implies that these sectors would gradually move into a more and more precarious economic position vis-a-vis younger and more efficient competitors in other countries. The relative difference observed depends on the longevity of capital goods as well as the rate of investment and the sectors studied have not been characterized by fast growth and/or a fast turnover of capital. However, if a large part of a country's industry is characterized by slow growth of investment that incorporates new techniques, structural problems and less future growth will result if the situation cannot be remedied, and the results suggest where further research should be most profitably directed.

Carlsson probes into the complex of factors influencing the choice of technology in an industry and the implications for industrial structure. Data obtained from direct interviews in U.S. and Swedish cement firms are used. Carlsson observes a strong relationship between relative factor prices on the one hand and the relative use of the same factors and the choice of technology in general on the other, if a long time perspective is allowed for. The other side of this is, of course, that sudden and strong relative price changes can cause sudden economic obsolescence in an industry which cannot adjust its production techniques fast enough. Another observation of interest is that ... technical change has gone hand in hand in the cement industry.

The conclusion is that even though one of the techniques studied is shown to be "theoretically" superior in every respect, its introduction was delayed, particularly in the United States, for at least a decade. Differences in relative factor prices (especially low energy prices in the U.S.) explain some of the delay, but it turns out that differences in market structure, raw material quality, past experience and attitudes as well as sloppy decision making also have played important roles.

The penultimate paper in the volume deals with production theory. *Färe* presents a new formulation of the law of diminishing returns within the dynamic production theoretic framework that he is currently developing together with Shephard. He is particularly interested in stating the conditions of when and how time availability bounds on essential factor inputs (like energy) also bound output over time. An essential input is one that causes output to be zero when the input level is reduced to zero. It is shown that in general there exist bounds on the time availability of essential inputs such that net output ceases before a finite horizon. Problems of this kind have very obvious practical applications, for instance in time scheduling of very complex production and assembly systems.

We do not believe that we have been able to handle the chosen problem to the full satisfaction of ourselves or others. The conclusion that emerges, however, is that the main obstacle to more knowledge and improved theory is lack of empirical information or facts. More empirical research and better techniques of measurement are the obvious priority and remedy.

Stockholm and New York, August 1978

Bo Carlsson *Gunnar Eliasson* *Ishaq Nadiri*

A VINTAGE MODEL OF SWEDISH ECONOMIC GROWTH FROM 1870 TO 1975

Ragnar Bentzel
IUI, Stockholm

I. INTRODUCTION

Econometric analysis of macroeconomic production functions has long been the standard method used in empirical studies of the casual factors behind the process of economic growth. The scientific literature is crowded with articles and books reporting different attempts to use such analysis for historical growth studies.[1] These attempts have, no doubt, made important contributions to our understanding of the growth process. There are, however, some weak points inherent in the production-function approach. A number of important features of the growth process cannot be analyzed because of the high level of aggregation. In addition, it is extremely difficult, not to say impossible, to construct reliable estimates of the capital-stock development, which is of fundamental importance for the analysis.

During the last twenty years, much attention has been paid to the vintage theory of capital, originally formulated and developed by Leif Johansen, Robert Solow and Edmund Phelps.[2] The essence of this theory is the assumption that capital of different ages is not fully malleable. This assumption implies, of course, that it is necessary

---

[1] Surveys of a number of different studies have been given by Centrum voor Economishe Studien [1974], Brown [1967] and Kennedy and Thirlwall [1973], amongst others.

[2] Johansen [1959], Solow [1960] and Phelps [1963].

to distinguish between amounts of capital that have been created at different points of time. By the introduction of this disaggregated way of looking at things, growth theory was enriched in several respects. In contrast to what is possible in an ordinary production-function model, a vintage model allows us to

(a) Make a distinction between embodied and disembodied, technological progress.

(b) Make a distinction between "ex ante substitutability" and "ex post substitutability" between labour and capital

(c) Treat capital scrapping as an endogenous variable, and

(d) Treat the time structure of investment as one of the determinants of the volume of production.

As an instrument of empirical analysis, the vintage approach has the very important advantage over the traditional production-function approach that it does not require capital-stock data. It is sufficient to have information about yearly investments. In those cases in which capital-stock data are not available, this advantage is, of course, decisive as regards the choice of approach.

In recent years, a number of studies have been made in which the vintage approach has been used for empirical analysis.[1] In most of these studies, the estimation of the rate of growth of technological progress has constituted the central point and in this respect some remarkable results have emerged. The models of the clay-clay type show, in general, a fairly high rate of growth of technological progress.[2] In contrast, the putty-clay models show a very low rate

---

[1] Bliss [1965], Attiyeh [1967], Baum, Görzig and Kirner [1971], Isard [1973], de Vries [1973/74], Benassy, Fouquet and Malgrange [1975], Görzig [1976], den Hartog and Tjang [1976], Kuipers and Bosch [1976], Sutton [1976] and Sandee [1976].

[2] Cf, for instance, den Hartog and Tjang [1976] and Benassy, Fouquet and Malgrange [1975].

of growth of such progress.[1] Furthermore, in those models which include not only embodied but also disembodied, technological – progress factors, the rate of growth of the embodied factor has turned out to be zero or very close to zero.[2]

Most of the empirical vintage studies that have been made so far have been attempts to find out the possibilities of using the vintage approach, in a fruitful way, for empirical analysis. As all these studies have been designed differently and for different purposes, it is difficult to give a general judgment as to whether the outcomes are to be regarded as positive or not. Some puzzling results have emerged and it is extremely difficult to make a fair appraisal of the realism of the models under consideration. It seems to be urgent to get more experience in this field of research.

The purpose of this paper is to report some additional experience of empirical analysis based on vintage models. For this purpose, I shall present a vintage model which I have constructed for the analysis of the economic development in Sweden from the beginning of the industrial revolution up to the 1970s. The general problem underlying the construction of this model can be formulated like this. Is it possible to construct a simple, one-sector, vintage model that is capable of simulating Swedish economic development during the period 1870-1975 and of giving non-trivial explanations for some of the characteristic features of the growth process during that period?

My model is, indeed, very simple. It includes only one sector – the whole Swedish economy, except public administration. Throughout the entire period under consideration, the economy is assumed to have been characterized by perfect competition and permanent equilibrium. In contrast to most other vintage models used for empirical analysis,

---

1 Cf Bliss [1965] and Görzig [1976].

2 Cf Bliss [1965], Isard [1973] and de Vries [1973/74].

it includes only one technological-progress factor, a labour-augmenting one. Other specific features are the assumptions that production within existing vintages decreases at a constant yearly rate and that the quantity of labour in existing vintages varies in inverse proportion to the labour-augmenting factor. The rate of interest plays a strategic rôle as a determinant of the life length of capital. Capital is scrapped for economic reasons only and at the point of time when labour costs tend to exceed the value of production. In new vintages, the volume of production is determined by a Cobb-Douglas function and there the labour share is constant. This implies that the capital-output ratio in new vintages is variable.

This procedure of parameter estimation differs radically from those used in earlier studies. The numerical specification of the model is given by using only information concerning the Swedish economy at the very beginning of the 1870s. Consequently, no information is used from the time-series which are to be explained.

The following presentation of my model is divided into four sections. The first one gives an account of how I have estimated the structure of the Swedish economy at the beginning of the period under consideration, i e in 1870. The second section gives a description of the model of the Swedish economy after 1870. The third section shows the results of the estimation of the development of the technological progress factor and, in addition, a simulation of the development of production and income distribution from 1870 to 1975. The fourth section, at last, gives some examples of concrete conclusions that can be drawn from a vintage model of the type presented in this paper.

II. THE ECONOMIC STRUCTURE OF SWEDEN AT THE BEGINNING OF THE 1870s

A necessary condition for the possibilities of using a vintage model for empirical analysis of the growth of an economy is that some basic facts are known concerning the structure of the economy in question at the beginning of the period under consideration. As my study covers the period from 1870 up to the present, the use of a vintage model for the analysis necessitated an attempt to estimate some characteristics of the Swedish economic structure at the very beginning of the 1870s. This attempt was made as follows.

The start of the industrial revolution in Sweden is commonly dated to the first few years of the 1870s. All empirical evidence shows that economic growth after the end of the 1860s became more rapid than it had been before. We do not know the growth rate at the beginning and the middle of the nineteenth century, since the Swedish national-income estimates do not go further back than 1860. However, the available figures of production in agriculture and the steel industry during the beginning and the middle of the nineteenth century indicate stationarity rather than growth in production per head. Since the population grew at a rate of 1 per cent per year during the pre-1870 period, I found it natural to assume that before 1870 the Swedish economy was characterized by a steady-state growth of 1 per cent per year.

For the further description of the initial structure, the following three basic assumptions were made:

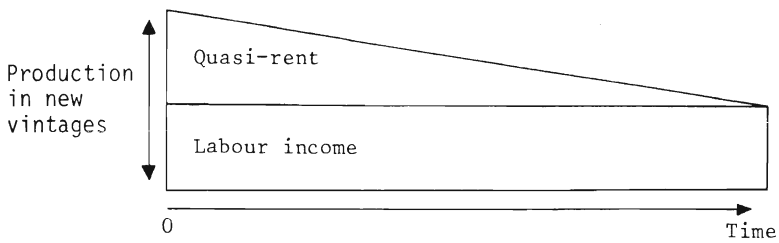
(a) The production volume associated with a certain vintage of capital was reduced - due to depreciation - by 1 per cent per year as time went on,

(b) Only those pieces of capital were used for which the value of production exceeded the labour costs, and

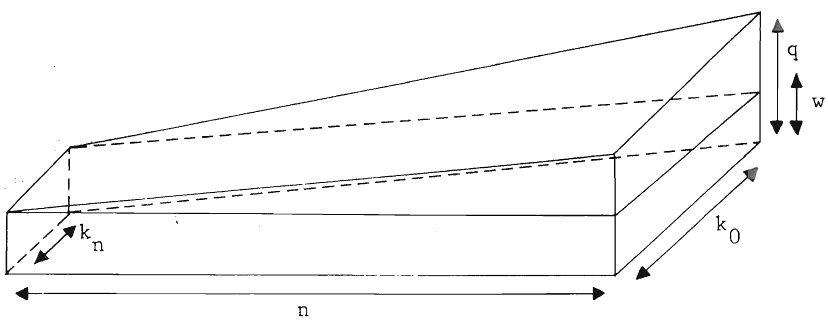
(c) Substitution between labour and capital was possible ex ante but not ex post,

(d) There was no technological progress.

On these assumptions, the development of production, the labour income, and the quasi-rent associated with a given amount of capital in period 0 can be illustrated like this:



Combined with the steady-state assumption made earlier, these three assumptions imply an economic structure that can be illustrated by a "box" of the following kind:



Here n illustrates the life length of capital and k_0 the volume of investment at the end of the period while k_n corresponds to the volume of investment n years earlier. The distance q shows the production per capital unit in a new vintage and the distance w represents the labour income per unit of capital.

In the following pages the following notations will be used:

q_{st} = Volume of production, associated with an s year old vintage in year t,

Q_t = Aggregated volume of production in year t,

ℓ_{st} = Number of employees associated with an s year old vintage in year t,
L_t = Total labour force in year t,
k_{0t} = Volume of investment in year t,
w_t = The real wage level in year t,
$(LW)_t$ = Total real labour income in year t,
n = The number of vintages in use,
α = The labour share of production in new vintages,
β = The rate of yearly decrease of production in existing vintages,
γ = The output-capital ratio in new vintages,
ε = The rate of steady-state growth before 1870,
V_{st} = The present value of the expected future profit stream associated with the s year old vintage in year t,
V_t = The sum of all V_{st} in year t,
r_t = The rate of interest in year t.

In accordance with the assumptions made above, the following equations will hold good

$$\alpha = e^{-n\beta}, \tag{1}$$

$$q_{0t} = \gamma k_{0t}, \tag{2}$$

$$k_{0,t-s} = k_{0t}e^{-\varepsilon s}, \tag{3}$$

$$Q_t = q_{0t}\underbrace{\int_0^n e^{-(\varepsilon+\beta)s}ds}_{N}, \tag{4}$$

$$(LW)_t = \alpha q_{0t}\underbrace{\int_0^n e^{-\varepsilon s}ds}_{M}, \tag{5}$$

where $\varepsilon = 0.01$ and $\beta = 0.01$.

According to the definitions of V_{st} and V_t we can, further, write

$$V_{st} = q_{st}\int_0^{n-s} e^{-(\beta+r)z}dz - 0.53q_{0,t-s}e^{-s\varepsilon}\int_0^{n-s} e^{-rz}dz \tag{6}$$

and

$$V_t = \int_0^n V_{st} ds. \tag{7}$$

Using equations (1), (2), (4) and (5) and denoting by M and N the two integrals appearing in (4) and (5), this system can be transformed into

$$\alpha = \frac{(LW)}{Q} \cdot \frac{M}{N}, \tag{8}$$

$$n = -(\log\alpha):\beta \quad \text{and} \tag{9}$$

$$\gamma = [CM]^{-1}, \tag{10}$$

where C stands for the investment ratio in the entire economy.

Since the "box" is meant to illustrate the Swedish economy at the end of the 1860s, these equations have to be consistent with the corresponding empirical data from that time. What matters in this context is that at the end of the 1860s the labour share of production, (LW):Q was 0.69 and the investment ratio, C, was 0.064. These values, inserted in the equations above, together with $\varepsilon = 0.01$ and $\beta = 0.01$, imply that[1]

$$\alpha = 0.53, \tag{11}$$

$$n = 63 \quad \text{and} \tag{12}$$

$$\gamma = 0.43. \tag{13}$$

These figures describe the "box" completely.

---

[1] Since the integrals M and N - after the numerical description of ε and β, - are functions of n only and the same is true of equation (9), we can solve the equations (8) and (9) for n and α.

The above assumption that the rate of yearly decrease of production within existing vintage amounts to 1.0 per cent is, in fact, not arbitrary. I shall now show that this value, in combination with the above values of n, α and γ, is consistent with the prevailing rate of interest. As shall be explained further in section IV, the rate of interest prevailing around 1870 can be estimated to 7 per cent, approximately.

From equation (6) can be concluded that

$$V_{0t} = 4.9q_{0t}. \tag{14}$$

Further it can easily be verified that

$$V_t = 2.8Q_t \tag{15}$$

or the equivalent value

$$V_t = 101q_{0t}. \tag{16}$$

The value V_{0t} consists of two parts, one corresponding to a net addition of capital amounting to 1 per cent of V_t and the other corresponding to the depreciation of the existing capital stock. Taking into consideration equations (14) and (16), it will easily be seen that these two parts amount to q_0 and $3.9q_0$. Consequently, the depreciation rate is 0.039.

As the Swedish economy before 1870 is assumed to have been stationary, the following relationship should hold good

$$(r+d)V_t = Q_t - (LW)_t,$$

where d is the depreciation rate. For $Q-LW = 0.31Q$ and $d = 0.039$, this equation gives

$r = 0.07$.

Consequently, the parameters calculated above are consistent with the empirical value of the rate of interest. As V_t is an increasing function of n, this condition of consistency will not be satisfied for other values of n.

In this context, it should be observed that V_0 is not identical with k_0. While k_0 is the value of investments in buildings, structures and machinery, V_0 includes in addition to these types of capital, also all other types of capital that are necessary for the production and marketing process, for instance, land growing forests, inventories, liquid assets, licences, etc.[1]

The quantity V_0-k_0 can, in fact, be interpreted in the following way: Suppose that the volume of production is determined by a production function $F(L,K,v)$ where $\underline{v}$ is the <u>volume of land</u>, inventories and other factors of production corresponding to V_0-k_0. Suppose further that the (L,K,v) combination chosen by the firms is determined by some profit maximization procedure. If only such optimal situations are considered the v-variable can be excluded from the production function, which accordingly can be written $H(L,K)$. Consequently, the existence of a difference between V_0 and k_0 is not a contradiction with the existence of an ordinary two-dimensional production function, provided that only optimal situations are considered.

In the following shall be assumed that the quantity V_0-k_0 has the character of fixed costs. Once invested it can never be regained. After the moment of investment the reward going to the factor of production v is therefore an inseparable part of the quasi-rent.

---

[1] According to the estimations above, V_0 is about twice as large as k_0. This does not seem to be too unrealistic. Old estimates of Sweden's national wealth indicate that, at the end of the nineteenth century, the value of natural resources and inventories was of the same order of magnitude as the total value of buildings, structures and machinery.

So far nothing has been said about the production functions of new vintages. This was not necessary for the description of the "box". In order to simplify the presentation in the next section, however, some remarks concerning the production function will be made here.

The production function in a new vintage will be assumed to be of the Cobb-Douglas type:

$$q_0 = A\, \ell_0^a\, k_0^b$$

where a+b=1. As the labour requirement is assumed not to change with the age of the vintage and the volume of production in existing vintages is assumed to be reduced by 1 per cent per year, the above description of the production function implies that the production in an s-year-old vintage can be written

$$q_s = A\, \ell_s^a\, h_s^b,$$

where $h_s=k_0 e^{-0.01s:b}$. By depreciating the capital in a proper way, we can, consequently, for all vintages, formulate a Cobb-Douglas production function with the same exponents as those appearing in the production function of the new vintage. This fact has the following implication. Let us suppose that the production function above holds good and let us define three aggregates L, K and Q in the following way:

$$L = \int_0^n \ell_s ds; \quad K = \int_0^n h_s ds \quad \text{and} \quad Q = \int_0^n q_s ds.$$

For given values of n, a and b, it is then possible to write

$$Q = B\, L^a K^b,$$

where B is a constant. This formula can now be used for determining the values of a and b in the following way.

The numerical description of the "box" implies that 1.1 per cent of the total employment and 0.78 per cent of the total production are associated with the oldest vintage. Let us suppose now that this vintage is scrapped. Since the two figures just mentioned can be identified with dL/L and dQ/Q, the following equation should hold good:

$$0.78 = 1.1a + (1-a)dK/K.$$

The total capital stock K is, of course, depending upon the rate of depreciation, which in its turn is determined by the labour elasticity of the production function. Furthermore, dK, i e the capital associated with the oldest vintage, is also determined by this elasticity. Consequently, dK/K, is a function of a only — for a given value of n — and the equation can be solved for a. The only value of a that satisfies the equation is

$$a = 0.6.$$

For the model construction in the next section, I have accepted this value and I have assumed that the production function elasticities remained constant and equal to 0.6 and 0.4 during the whole period up to 1975.

III. THE MODEL OF SWEDISH GROWTH SINCE 1870

The model described in the preceding section refers to a steady-state growth with no technological progress. In the following pages, it will be called the "stationary model". In this section, I shall give an account of the more general model, which I have constructed for the analysis of Sweden's economic growth in modern times, here defined as the period 1870-1975. This model will be called the "growth-model".

In the construction of the growth model, I maintained the stationary model as a skeleton, so that the former can be regarded as a modified version of the latter. The modifications are, however, quite essential A growth-creating, technological-progress factor has been introduced and the following parameters appearing in the stationary model have been made variable: the life length of capital, the capital-output ratio in new vintages, the capital intensity in new vintages and the rate of production depreciation within existing vintages.

The technological-progress factor

Only one single kind of technological-progress factor is introduced in the model, a disembodied, labour-augmenting factor. The motives for choosing this and only this progress factor were briefly the following:

Experiments with different combinations of labour- and capital-related factors and with different combinations of embodied and disembodied factors yielded clear and uniform results. They all indicated that the disembodied, labour-augmenting factor was greatly predominant. When included in the model, the other types of progress factors had only small effects on production and, in addition, they behaved "irrationally", in the sense that they showed unexplainable ups and downs with no systematic trends. This experience is in good accordance with the above-mentioned results of those earlier studies in which both embodied and disembodied progress factors were included.

The predominance of the disembodied, labour-augmenting factor can be explained also by a more general consideration. Looking at the sta-

[1] See, for instance, Bliss [1965], de Vries [1973] and Isard [1973].

tionary model, it is easy to conclude that a wage increase implies one of two alternative types of change, either a decrease in the number of vintages in use or a productive gain in the oldest vintage. The first of these two alternatives cannot, alone, give rise to more than a very modest, long-run, wage growth without leading to an unreasonably large decrease in the number of vintages. The second alternative must imply the existence of disembodied, technological progress, either labour-augmenting or capital-augmenting. However, from a glance at the empirical data of employment, wages and capital formation, it is easy to conclude that the capital-augmenting factor, if present, cannot have been very important. The reason is that the combination of an even rather small, capital-augmenting factor and such a fast-growing, capital formation as occurred in Sweden at the end of the nineteenth century would imply a much higher rate of employment growth than the actual one. The general conclusion to be drawn from these facts is, of course, that the only technological-progress factor that — within the framework of my model — can give a reasonably good explanation of the Swedish wage growth after 1870 is a disembodied, labour-augmenting factor.

Since embodied, technological-progress factors cannot create wage increases in the old vintages, the assumption that all technological progress is of an embodied character cannot be consistent with a rapid wage growth. Such an assumption is, in addition, inconsistent with the available data also in another respect. In my model, the conditions of equilibrium in the new vintages would imply that a long-term increase in embodied, technological progress should result either in a downward trend in the price ratio between capital goods and consumer goods or in an upward trend in the cost of capital. However, the Swedish data do not show such trends.[1]

The way in which a labour-augmenting factor should be introduced into the model was fairly self-evident. Taking the stationary model

---

[1] I cannot, of course, deny the existence of capital augmenting technological progress. The fact that they are difficult to discern, statistically, is perhaps due to the existence of one or more neutralizing factors, for instance the gradual reduction of capital utilization caused by the shortening of the time of work.

as a point of departure, we can denote by ℓ_{st} the labour quantity associated with an s-year-old vintage in year t. In the growth model this variable was quite simply replaced by the variable $\ell_{st}x_t$, where x_t ($x_0 = 1.0$) denotes the accumulated value of the technological-progress factor from 1870 (t=0) up to the point of time t. This variable x has, obviously, the character of a labour-efficiency factor and in the following pages, the ratio w_t/w_0x_t, where w_0 is the wage level in 1870, will be called the wage-efficiency factor. It will be denoted by y_t.

After the introduction of the x-factor, the production function in new vintages will be

$$q_{0t} = A(\ell_{0t}x_t)^{0.6}k_{0t}^{0.4}. \qquad (17)$$

Since the x-factor in this equation can be put outside the bracket, it cannot be identified as a labour-augmenting factor. What makes such an identification possible is the assumption that this x-factor affects also the labour requirement of existing vintages. More precisely, it is assumed that the volume of labour associated with an s-year-old vintage in year t is

$$\ell_{st} = \ell_{0,t-s}x_t^{-1}x_{t-s}, \qquad (18)$$

a formula which implies that in existing vintages the labour quantity is gradually reduced at the same rate as the technological-progress factor x is increasing. Consequently, an increase in the x-factor of z per cent implies a decrease of z per cent in the labour input in all existing vintages.

The labour share

In the stationary model, the labour share in new vintages was estimated as 0.53. But how should it be assumed to vary in the growth model? As a basis for my consideration of this question, I took the well-known fact that in most countries the labour share of total production has remained fairly stable. This fact indicates a long-run stability of the labour share in new vintages.[1] So I have made the very simple assumption that the labour share in new vintages remained constant during the whole period 1870-1975. This assumption means that

$$w_t \ell_{st} = 0.53 q_{0,t-s} y_t y_{t-s}^{-1}. \tag{19}$$

The capital-output ratio

The assumption of a constant labour share has an immediate implication for the capital-output ratio on new vintages. By substituting $0.53q/w$ for ℓ in the production function formula (17) we get, after some manipulations, the following equation:

$$q_{0t}/k_{0t} = B(x_t/w_t)^{1.5},$$

where B is a constant. With the above definition of the variable y, this equation can also be written

$$q_{0t}/k_{0t} = B' y_t^{-1.5}, \tag{20}$$

which shows that the output-capital ratio is proportional to the 1.5 power of the inverted, wage-efficiency ratio.

The production-depreciation factor

In the stationary model, it was assumed that production within each

---

[1] It should be observed that the constancy of the labour share does not follow from the constant elasticity property of the Cobb-Douglas function. The reason is that production decreases as time goes on. In fact, the present value of the expected stream of quasi-rents coming from a new investment project can be written as

$$q_0 \int_0^n e^{-(\beta+r)s} ds - w\ell \int_0^n e^{-rs} ds.$$

By maximizing this expression we get

$$\frac{w\ell}{q} = a I_1 I_2^{-1}$$

where I_1 and I_2 are the two integrals above and a is the labour elasticity parameter in the Cobb-Douglas function.

existing vintage was reduced by 1 per cent per year. A similar depreciation factor is assumed to exist in the growth model, but there it is supposed to be variable. Accordingly, the production in an s-year-old vintage can be written

$$q_{st} = q_{0,t-s}e^{-s\beta_{t-s}} \tag{21}$$

where β_{t-s} is the depreciation factor associated with capital invested in $t-s$.

On the assumption of static expectations, the consistency of the model implies that a decrease of the life length of capital is followed by an increase of the production depreciation rate in the future vintages.[1] With a constant labour share of 0.53, the following equation has to be satisfied:

$$e^{-n_t\beta_t} = 0.53,$$

which implies that

$$\beta_t = 0.63n_t^{-1}, \tag{22}$$

where the index t refers to the period of time when the vintage was "born".

Number of vintages

A central feature of the model is the assumption that only those vintages are used in which the value of production is not less than

---

[1] It should be observed that a change in β can occur only simultaneously with a change in the capital intensity in new vintages. On the assumption that there is a relationship between the capital intensity and the costs of repair and maintenance, it is obviously possible to interpret an increase in β as a consequence of an increase in the repair and maintenance expenditures caused by the change in capital intensity. On this assumption, it is, furthermore, possible to imagine a profit-maximization procedure, by which the labour share and the production-depreciation factor β are determined simultaneously.

the labour costs. This assumption implies, of course, that the value of production in the oldest vintage equals labour costs, and that a wage rise is possible only if either the labour-augmenting factor rises or the number of vintages is reduced. In the former case, the wage level can rise in the same proportion as the productivity factor. In the latter case, every year of decrease in the life length of capital gives room for 100 × β per cent increase in the wage-efficiency ratio. Consequently, for all years in which the scrapping refers to vintages in which the production-depreciation rate is 0.01 we can write

$$(63-n_t)0.01 = \frac{w_t}{w_0 x_t} - 1$$

or

$$1 + (63-n_t)0.01 = y_t. \tag{23}$$

For years in which the scrapping refers to vintages in which the production depreciation factor differs from 0.01 the corresponding equation can be written

$$\beta_{t-n}\Delta n_t = \Delta y_t. \tag{24}$$

In the analysis below it so happens that all scrapping refers to vintages with a depreciation factor of 0.01 except the scrapping during the 1970s. This means that the equation (23) is valid for all years up to 1970 and the equation (24) refers to the years after 1970 only.

The rate of interest

The assumption of perfect competition implies that the discounted value of the expected income stream of quasi-rents emanating from a new investment project should equal the total investment costs, V_0. Consequently, the following equation should hold good:

$$q_{0t}I_1 - w_t \ell_{0t} I_2 = V_{0t}, \tag{25}$$

where the I:s are defined as

$$I_1 = \int_0^n e^{-(\beta_t + r_t)z} dz \quad \text{and} \quad I_2 = \int_0^n e^{-r_t z} dz.$$

Since a wage increase proportionate to a corresponding increase in the productivity factor leaves labour income and production value unchanged, such a wage change will not affect the variables in the equation above. The situation is, however, different for a change in the wage-efficiency ratio. If the rate of return of the investment project is not to be worsened by a rise in the wage-efficiency ratio the rate of interest must fall so much that the labour-cost increase is compensated by a decrease in capital costs. Consequently, there must be a relationship between the wage-efficiency ratio and the rate of interest.

In the preceding section, it was shown - equations (13) and (14) - that the stationary model implied that

$$V_{0t} = 4.9q_{0t} \quad \text{and} \quad k_{0t} = q_{0t}/0.43 = 2.3\, q_{0t},$$

which in turn implies that

$$V_{0t} = k_{0t} + 2.6q_{0t}. \tag{26}$$

This equation is assumed to hold good also for the periods after 1870, an assumption which implies that the value of capital not included in the figures of investment, i e land, inventories, etc, varies in proportion to the volume of production in new vintages.

Inserting the right-hand member of equation (26) in equation (25), we get

$$I_1 - 0.53I_2 = k_{0t}/q_{0t} + 2.6, \tag{27}$$

an equation which includes four variables, n, β, k_0/q_0 and r.

Since β and k_0/q_0 are uniquely determined by n, according to equations (22), (20) and (23), we can regard (27) as an equation between n and r only. Given r, we can consequently determine n, and vice versa. Therefore, we can formally write equation (27) as

$$F(n_t, r_t) = 0. \tag{28}$$

The mechanism behind this equation obviously means that the rate of interest and the wage-efficiency ratio act as two communicating vessels. If the wage-efficiency ratio is raised, the rate of interest must fall. If not, investment projects will show expected losses and therefore no investment will take place.

The model equations

By bringing together equations (28), (23), (24), (21), (20), (22), (17), (18) and (19), we get the following complete description of the growth model:

$$F(n_t, r_t) = 0, \tag{29}$$

$$y_t = 1 + 0.01(63 - n_t) \text{ for all years before 1870,} \tag{30a}$$

$$\Delta y_t = \beta_{t-n} \Delta n_t \text{ for the 1970s,} \tag{30b}$$

$$\beta_t = 0.63 n_t^{-1}, \tag{31}$$

$$q_{0t} = A(\ell_{0t} x_t)^{0.6} (k_{0t})^{0.4}, \tag{32}$$

$$q_{0t} = B y_{0t}^{-1.5} k_{0t}, \tag{33}$$

$$q_{st} = q_{0,t-s} e^{-\beta_{t-s}}, \tag{34}$$

$$\ell_{st} = \ell_{0,t-s} x_t^{-1} x_{t-s}, \tag{35}$$

$$w_t \ell_{st} = 0.53 q_{0,t-s} y_t y_{t-s}^{-1}. \tag{36}$$

By simple summation, we can, of course, also form the three aggregates

$$Q_t = \int_0^n q_{st}, \quad (LW)_t = \int_0^n (w_t \ell_{st}) \text{ and } L_t = \int_0^n \ell_{st}. \tag{37}$$

Furthermore, by using equation (23) we can determine the labour-augmenting factor like this:

$$x_t = (LW)_t (LW)_0^{-1} L_0 L_t^{-1} [1 + 0.01(63-n_t)]^{-1}. \tag{38}$$

A glance at the above equation system indicates that, given the time-series of the investment volume and the interest rate, equations (29)-(34) make it possible to determine, in turn, the variables n_t, y_t, q_{0t}, β_t, and q_{st}. Consequently, the aggregated production Q_t can also be determined. Furthermore, the values of q_{0t} and y_t can be used to determine $w_t \ell_{st}$ by equation (35) and consequently the aggegated labour income $(LW)_t$ can also be obtained. All this together means that access to empirical data showing the time-series of the volume of investment and the rate of interest enables us to simulate the corresponding time-series of total production and total labour income. Access to data on total employment enables us, in addition, to simulate the development of the labour-augmenting factor x_t. These properties of the model have been used for the simulation procedure that will be described in the next section.

The propelling factor of the "model economy" is assumed to be the labour-augmenting factor x. The time path of this factor is regarded as exogenously given. When it grows, it creates disequilibrium tendencies which put the whole system into motion.

In the very long run, total employment must, reasonably, develop close to the total labour force. Therefore, my model makes no distinction between these two variables. They are assumed to have identical values. However, a conceptual distinction should nevertheless be made, because the total labour force has to be regarded as exogenously given, while the total employment is determined as an endogenous variable in the model. In fact, total employment should be looked upon as a target variable determined - either by a labour-market mechanism or by economic-policy measures - in such a way that it will equal the total labour force.

There are two more variables whose status in the model has not been made clear - the rate of interest and the volume of investment. As regards their character of exogenous or endogenous variables, different interpretations are possible. One alternative is to regard the rate of interest as exogenously given. The consistency of the model requires in this case that the volume of investment is determined - either via a wage policy or via some investment affecting government policy - in such a way that full employment is attained. Another alternative is to regard the volume of investment as exogenously given and to regard the rate of interest as a policy parameter, used as an instrument for attaining full employment. Yet another alternative is to regard the wage-efficiency ratio as given by the labour-market mechanism and to regard the rate of interest and the volume of investment as policy parameters, used for creating equilibrium and full employment.

The fact that the model allows for different interpretations of the casual order does not, of course, mean that one of these alternatives is to be regarded as the right one and the others as wrong. It is, in fact, quite possible to imagine that the different alternatives refer to different periods of time. Furthermore, it should be observed that the simulation results are independent of the choice of alternative.

IV. THE SIMULATION RESULTS

In the preceding section, I showed that access to time-series of the volume of investment and the rate of interest makes possible a simulation of all the relevant variables included in the model. This property of the model has been used for the simulation procedure to be reported in this section, together with the simulation results. This procedure is in fact very simple.

According to the model, the rate of interest determines uniquely the number of vintages and the output-capital ratio in new vintages. This means that, starting from the year 1870, we can gradually estimate (period by period), the total production and the total labour income by the following two equations:

$$Q_{t+1} = \gamma_{t+1} i_{t+1}(1+\beta_{t+1})^{-1/2} - q^{s}_{t+1} - \sum_{s=t_{0,t+1}}^{t} q_{st}(1+\beta_{t-s})^{-1} \text{ and} \tag{39}$$

$$(LW)_{t+1} = 0.53\ \gamma_{t+s} i_{t+1} - (\ell w)^{s}_{t+s} - [(LW)_{t} - (\ell w)^{s}_{t+1}](1+\beta_{t+1}\Delta n), \tag{40}$$

where γ denotes the output-capital ratio in new vintages and i the volume of investment. The variables q^s and $(\ell w)^s$ stand for the volume of production and the labour income, respectively, in vintages scrapped during the period. The symbol β_s is the production depreciation factor, referring to the vintage invested in s, Δn is the decrease in the number of vintages under the period and $t_{0,t+1}$ is the period of time to which the oldest vintage refers.

Knowing the development of Q and (LW) up to the point of time t and in addition, the values of i_{t+1} and r_{t+1}, all the terms in the right-hand members of equations (39) and (40) can be determined and, consequently, also the left-hand members.

In order to simplify the calculations, I have used throughout 5-year averages of the investment figures. This means that the value of β in the equations above has to be thought of as being approximately five times as high as its 1-year equivalent. It should be observed that the values of Q and LW, which emerge from the simulations, refer to separate years, not to 5-years averages.

For the simulation procedure and for the comparison between simulated and actual values, the following four time-series were needed:

(1) the volume of production in the private sector of the Swedish economy, (2) the volume of investment in this sector, (3) the labour share of production in this sector and (4) the rate of interest (or, more correctly, the cost of capital). The first two of these time-series could easily be constructed by some minor manipulations with data published elsewhere.[1] For the post-war period, the desired income-distribution figures have been provided by the Swedish Employers' Confederation.[2] For the period before 1950, new data were constructed by making some modifications to the data presented in an earlier study.[3]

The estimation of a time-series showing the development of the rate of interest was a little problematic. For the period before the First World War, the statistical information about different rates of interest is very incomplete. However, it can be concluded that the interest rates of industrial bonds issued by big firms varied between 5 and 6 per cent and that the bank rates were 1 or 2 per cent higher. These rates remained at the same level, approximately, during the 1920s, but at the beginning of the 1930s, there was a sudden fall by a couple of percentage units. With the exception of the war years, this low rate was maintained until the middle of the 1950s, and since then the nominal rates of interest have been higher. However, the real rates - which seem to be the relevant ones in this context - have remained very low, about 3 per cent as an average, for the 1950s and 1960s. Since 1970, the real rate has been approximately zero.

In the study presented in this paper, there seems to be little sense

---

[1] Krantz and Nilsson [1975] and National Accounts.

[2] The figures for the after-war period shown in table 1 on p 39 are 3 percentage units lower than the corresponding figures given by the Swedish Employers' Confederation. This is due to the fact that my figures had to be chained to the series for the period before 1950. Consequently my figures are probably 3 percentage units too low.

[3] Jungenfelt [1966].

in using sophisticated methods of determining the year-to-year development of the rate of interest. Instead, an extremely schematic procedure has been chosen. For the simulation, I have quite simply allowed for a constant rate of interest of 7 per cent all the time from 1870 to 1930 and a rate of 5 per cent from 1930 to 1950. For the period 1950 to 1970, I have allowed for 3 per cent and for the first part of the 1970s for 0 per cent.

The growth path in the efficiency factor is estimated by the quantities of labour measured by the number of individuals. From many points of view, it might have been better to proceed not from the number of individuals but rather from the number of working hours. As the data are lacking for earlier periods, it has not been possible to do it in this way without a loss of comparability between periods. Those who want to relate the efficiency factor to working hours instead of individuals can easily do so. It is only necessary to add to the estimated value of growth in the efficiency factor the growth of the ratio of the number of individuals employed to the number of hours worked. From 1950 to 1972 this ratio has grown by 0.15 per cent per year on the average.

The results of the simulation are shown in Tables 1 and 2. They can be summarized like this:

(1) In view of the very long period covered by the simulation and of the fact that the simulation has been performed without using information from the time-series to be explained, the conformity between the hypothetical and the actual values seems to be remarkably good. This good fit justifies a positive answer to the first part of the basic problem raised in the introductory section. There it was asked whether it is possible to construct a simple, one-sector model that is capable of making possible a close-to-reality simulation of Swedish economic development during 100 years. The figures presented in Table 1 confirm this possibility.

(2) The good fit between the simulated and the actual values supports the general hypotheses underlying the model, including the hypothesis that the technological progress has been predominantly disembodied and labour-augmenting.

(3) The simulation indicates that the lifetime of capital was constant during the first 60 years of the period under consideration and that it fell thereafter to 40 years in 1970 and to 30 years in 1975. This fall in the number of vintages is in agreement with the results of some other studies.[1]

(4) According to the simulation, the output-capital ratio decreased from 0.43 during the period 1870-1930 to 0.26 at the beginning of the 1970s. Simultaneously, there was a gradual increase in the ratio of capital depreciation to gross investment. The same type of development has been found in other studies.[1]

(5) It must be admitted that the realism of the assumption made above concerning the relationship between the rate of interest and the number of vintages - equation (28) - is doubtful. Therefore, it may be worth while to investigate the consequence of giving up that assumption. This can be done by estimating the number of vintages, on the assumption that the simulated and the actual values of aggregate production coincide during the whole period. The result of this calculation was as follows:

Year	1870	1890	1910	1930	1950	1955	1960	1965	1970	1975
Number of vintages	63	60	60	64	47	42	39	40	35	31

A comparison with the figures given in Table 1 shows that the series in question are nearly identical except for one single year, 1970. This indicates that the assumed relationship between the rate of interest and the number of vintages is in good agreement

---

1 Cf, for instance, den Hartog & Tjang [1976].

Table 1. Estimations of production, labour income, labour share, output-capital ratio and number of vintages

	1890	1910	1930	1950	1955	1960	1965	1970	1975
Production									
Actual (1870=100)	167	322	602	1014	1107	1307	1669	1949	2157
Simulated	171	326	594	1021	1099	1333	1670	2042	2155
Error margin, %	+2.4	+1.2	-1.3	+0.7	-0.7	+2.0	-0.1	+4.8	0.0
Labour income									
Actual (1870=100)	163	299	549	940	1122	1325	1716	1948	2218
Simulated	163	305	556	1009	1137	1336	1640	1987	2271
Error margin, %	0.0	+2.0	+1.3	+7.3	+1.3	+0.8	-4.6	+2.0	+2.4
Labour share[a]									
Actual (1870=100)	0.67	0.64	0.63	0.64	0.70	0.70	0.71	0.69	0.71
Simulated	0.66	0.65	0.65	0.68	0.71	0.69	0.68	0.67	0.73
Estimated number of vintages	63	63	63	49	40	40	40	40	30
Estimated output-capital ratio	0.43	0.43	0.43	0.34	0.30	0.30	0.30	0.30	0.26

[a] Cf note No. 2 on p. 36.

Table 2. Estimations of yearly growth rates and the yearly growth of technological progress

	1870-1890	1890-1910	1910-1930	1930-1950	1950-1955	1955-1960	1960-1965	1965-1970	1970-1975
Yearly growth rates, %									
Actual	2.6	3.3	3.1	2.6	1.8	3.3	4.9	3.1	2.0
Estimated	2.7	3.2	3.0	2.7	1.5	3.9	4.5	4.0	1.1
Estimated yearly growth of technological progress	1.9	2.5	2.1	1.0	1.6	3.1	5.4	2.7	0.7

with the other assumption of the model.

(6) The error margins presented in Table 1 are in most cases small. There are, however, three exceptions. They refer to labour income in 1950 and 1965 and to production in 1970. It is not very easy to understand why the simulation gives such a bad fit for the labour income of 1950 and 1965. The bad fit for production in 1970 can, however, easily be explained. The capital costs for Swedish industry were, no doubt, lowered during the latter part of the 1960s by a number of economic-political measures aimed at the stimulation of investments; the investment funds were released much more generously than previously and large subsidies were given to firms starting new plants in backward areas. It seems, in fact, that the assumption of a 3 per cent rate of interest during this period is not very realistic. The large margin of error in Table 1 and the figure given for 1970 under paragraph (5) above indicate strongly that there was a decrease in the number of vintages by about 5 during the period 1965-70.

(7) The rate of growth of the labour-augmenting factor has varied around a value slightly above 2 per cent per year, which seems to be a "normal value". That the rate was higher during the period 1890-1910 is not surprising, if we consider the exceptionally good conditions for economic growth that pertained during that period. Nor is it surprising that the rate was exceptionally low during the period 1930-50. The high rate 1965-70 and the low rate 1970-75 can be explained by what was said above, namely, that a part of the estimated decrease in the number of vintages for the period 1970-75 in reality occurred already during the end of the 1960s; the average of the growth rate for the 10-year period 1965-75 was 1.7 per cent. Also for the two periods of the fifties the average was rather normal. The low rate at the beginning of the 1950s and the high rate at the beginning of the 1960s do not, however, fit into the "normal" picture.

(8) The estimated values of the rates of growth of the labour-augmenting factor agree rather well with the estimate made in an earlier

Swedish study using a production-function approach.[1] The disembodied technological factor — divided by the labour elasticity in order to be comparable with a labour-augmenting factor — was estimated to have been 2.2 for the period 1870-1964. The figures in Table 2 are also in a rather good agreement with the results obtained by C E Ferguson and P A David and Th van de Klundert in aggregated production-function studies of the U S economy.[2] Ferguson's analysis yielded a labour-augmenting factor of 1.9 for the period 1948-63, while David's and van de Klundert's investigation, which covered the period 1899-1960, indicated a labour-augmenting factor of 2.3.

---

[1] Y Åberg [1969].

[2] C E Ferguson [1965] and P A David and Th van de Klundert [1965]

V. THE EXPLICATORY POWER OF THE MODEL

The scientific value of a model of the above type is, of course, dependent on the possibilities of using it for drawing concrete conclusions concerning reality. In making a general appraisal of the model, it is, consequently, important to get some information about its power to explain actual economic phenomena. The purpose of this section is to give some information of that kind, by presenting some examples of conclusions that can be drawn from the model presented in the preceding section. These examples refer, of course, to Swedish development, but it should be borne in mind that my purpose is not to present an analysis of the Swedish growth process but only to show that a very simple, one-sector, vintage model may allow us to draw some important conclusions.

As will be seen from Table 1, the growth rate of the Swedish economy has varied from one period to another. Most of these variations have been simulated correctly by the model and, in that sense, the simulation can be said to explain the variations in the rate of growth. This is true also for the period of high growth-rate between 1890 and 1910 and the extreme boom period of 1960-65. According to the model, the production increase during these periods was caused by the high investment ratio. Also the slow rate of growth at the beginning of the 1950s is fairly well mirrored by the simulation. The slow growth during these years is explained by the model by the extra scrapping that occurred as a consequence of an increase in the wage-efficiency ratio.

It is certainly true that the extreme boom during the first half of the 1960s does not give rise to "difficulties of explanation" if we look only at the production side of the model. However, if we look at the labour side, such difficulties will arise. The problem is how all the new, invested capital could be manned without pulling more than the "normal" amount of labour from the oldest vintages. According to the model, this was possible because of a sudden jump in the labour-augmenting factor. But why did this jump happen? The model cannot, of course, give an answer to that question, but it has raised the problem.

Within the framework of the model, it is hardly meaningful to disaggregate the growth of production into parts interpreted as separate effects of changes in capital stock, employment and technological progress. However, the model does allow of assessments of the marginal productivity of capital and labour. For labour, such an assessment is trivial. For capital, it is not so. It is, in fact, possible to estimate not only the marginal productivity that is of relevance to the private investor but also the social, marginal productivity, defined as the increment in total production in consequence of an increase in investments at a constant level of employment. Of course, such a change implies a transfer of labour from the oldest to the newest vintages. Estimates of the social, marginal productivity defined in this way indicate that it amounted to 20 per cent during the period 1870-1930. After 1930, it decreased and in 1975 it was no more than 12 per cent.[1]

---

[1] On the assumption that the initial situation is characterized by full employment, the production increase per unit of incremental capital can be written

$$dQ/dk_0 = (dq_0 - dq_n) : dk_0,$$

where dq_n stands for the production in the vintage, scrapped because of the necessary transfer of labour to the extra new capital. For the period 1870-1930, the output-capital ratio in new vintages remained constant and equal to 0.43. During that period, the ratio between the labour productivity in the oldest vintage and the productivity in the newest vintage was 0.53. Consequently, the derivate dQ/dk_0 is equal to

$$0.43(1-0.53) = 0.20.$$

In 1975, the output-capital ratio in new vintages was 0.26. This implies that dQ/dk_0 for 1975 was 0.12.

The marginal productivity of capital was defined as the ratio between the increment of production in year t, following from the hypothetical extra investment at the beginning of that year and the volume of this extra investment. However, investments in year t affect production also in the years t+1, t+2, etc. If the entire series of consequential increments to production is known - net after deduction of the corresponding production loss in the oldest vintage - it is, of course, possible to estimate the social rate of return of the extra investment. Such an estimate shows that the internal rate of return, according to the model, amounted to 18 per cent until 1930 and thereafter decreased to less than 10 per cent in 1970.

The long-term development of the Swedish functional distribution of income is characterized by a reduction in the labour share from 1870 to 1930 and by two, sudden, upward jumps of the labour share, one at the beginning of the 1950s and one at the beginning of the 1970s. In "the world of the model", the reduction in the labour share until 1930 is explained by the combination of an unchanged number of vintages and a shift in the centre of gravity of the production structure towards younger vintages, where the labour share is lower than in the older ones. The jumps at the beginning of the 1950s and the 1970s are explained by the decrease in the number of vintages. A decrease in the number of vintages implies a tendency to raise the labour share.

The combination of an acceleration of the investment growth and a non-decreasing number of vintages implies, in the "world of the model", a decrease in the labour share of production.[1] If this mechanism is realistic, it has an important consequence for economies that are at the beginning of the industrialization process and have an abundant labour supply. On the traditional assumption that the saving rate from capital incomes is higher than that from labour incomes,

---

[1] J Sutton [1976] deals fairly much with this mechanism. He shows that the combination of an investment acceleration and an elastic labour supply results in a lowering of the labour share. He explains the development in Japan by this mechanism.

the income redistribution caused by an investment acceleration creates automatically at least some of the additional saving that is needed for financing the investment growth. In Sweden, this savings-creating mechanism seems to have been very important, especially during the period 1890-1910.

The model indicates that the number of capital vintages was constant during the entire period of 60 years from 1870 to 1930. This constancy implies that the wage rate increased at the same rate as the labour-augmenting factor, which in turn means that the labour costs remained constant. Since the rate of interest did not change very much during this period, there were no incentives to substitute capital for labour — or vice versa — during this period. It was, according to the model, not until the depression during the 1930s that substitution started to take place. The fall in the rate of interest provided incentives to use more capital-intensive methods of production than before.

According to the model, the labour productivity is higher in new vintages than in the older ones. This means that the ratio between total production and total labour force is influenced by the vintage structure; the larger the young vintages, the greater is the aggregated productivity. This property of the model is important as regards the problem of estimating the productivity gains attained by the transfer of labour from agriculture to industry. According to the actual model, a great part of the productivity gap between manufacturing industry and agriculture that existed in Sweden up to the Second World War can be explained quite simply by the difference in the vintage structure between the two sectors. The labour productivity was higher in manufacturing industry than in agriculture, because the mean age of capital was lower in the former sector than in the latter. This does not, of course, imply a difference in marginal productivity between the two sectors.

At the beginning of the 1930s, there was obviously some type of structural shift in the Swedish economy, a shift from a situation characterized by unaltered labour costs (unaltered for augmented labour), lack of substitution between labour and capital and a downward long-term trend in the labour share of production to a situation characterized by increasing labour costs, substitution between labour and capital and an increasing trend in the labour share. In trying to find the explanation of this shift, we immediately encounter the problem touched upon in section III, viz. how to interpret the casual order of the model. There are, in principle, two different alternatives to choose between.

As I stated earlier, one way of looking at the causal order is to regard the rate of interest as an exogenous and casual factor. This implies that the casual order can be thought of as follows. On account of the fall in the rate of interest, the capital costs in new vintages decreased, which created room for an increase of the wage-efficiency ratio in the new vintages. This increase was spread over the entire labour market and forced an extra amount of scrapping of old vintages, which in turn produced a tendency to unemployment. This tendency was, however, never realized, because the lowering of the rate of interest stimulated investments enough to make it possible for the labour freed by the extra scrapping of old capital to be absorbed by the manning of new capital.

The other interpretation alternative is to consider the rise in the wage-efficiency ratio as exogenous and to regard the structural shift as an effect of institutional changes caused, for instance, by a transition from one type of economic policy to another, from one labour-market mechanism to another, etc. One can, for example, imagine an institutional change leading to increased wage pressure, which forces the authorities to lower capital costs in order to compensate for increased labour costs and to avoid the unemployment tendencies arising from the increased scrapping of old capital.

In the Swedish economy, there has been a substantial increase in the ratio of capital depreciation to gross investment. This development is fairly well mirrored by the model. In the "world of the model", the ratio in question increased from a low of less than 40 per cent in 1950 to around 65 per cent at the beginning of the 1970s. The explanation of this development is the increase in the frequency of vintages with high production-depreciation rates.

An increase in the ratio of capital depreciation to gross investment means, of course, a tendency to a lower growth rate, given the volume of investment. Therefore, the development mentioned in the preceding paragraph has meant a lowering of the growth potential of the Swedish economy. Earlier in this section, I argued that this potential was impaired also by another phenomenon, the decline in the output-capital ratio. Consequently, there are at least two factors that create important tendencies to worsen the growth potential of the Swedish economy. The model indicates that these tendencies started to assert themselves in the middle of the 1930s and that they have grown in strength, especially since the middle of the 1960s.

The appearance of the growth-potential-worsening factors mentioned in the preceding paragraph is, in the model, a consequence of the decrease in the number of vintages. This decrease in its turn is a consequence of the high investment level; the manning of all new capital necessitated the pulling of labour from the oldest vintages. If this mechanism has a general validity, it implies that the possibilities of promoting growth in a full-employment society by expanding investments are narrowly limited. The more investments are expanded, the more the growth-counteracting factors will worsen the growth potential. This conclusion is certainly in full agreement with the traditional assumption of the decreasing marginal productivity of capital, but in the model presented above, this marginal-productivity effect is reinforced by others working in the same direction.

It is well known that a traditional production-function model can be used for forecasting future production for given values of the volume of investments, the volume of labour and the productivity factor(s). The same types of forecasts can be made with the aid of a vintage model of the type presented in this paper. My model has, in fact, been used for a number of such estimates. All these estimates have shown that - given a normal 2-per-cent increase of the labour-augmenting factor - an extreme increase in the investment ratio will be necessary, if the Swedish economy is to be able to attain a growth rate of 3 per cent per year or more. This means a much lower growth potential than before. The reasons are, of course, those mentioned above - the decrease in the output-capital ratio, and the higher rate of capital depreciation.

In the introductory section was stated that the general problem underlying the construction of the model presented in this paper was to find out whether it is possible to construct a simple one-sector model that is capable of simulating the Swedish economic development during the last one hundred-year period and of giving non trivial explanations for some of the characteristic features of the growth process during that period. The first part of this problem was answered positively in the preceding section. The discussion in this section has shown that the model has a good capability of explaining specific features of the growth process and that, consequently, also the second part of the above problem can be answered in the affirmative.

REFERENCES

Attiyeh, R., 1967, Estimations of a Fixed Coefficients Model of Production. Yale Economics Essays 7 (Spring 1967) 1-40.

Baum, D., Görzig, B. and Kirner, W., 1971, Ein Vintage-Capital-Produktionsmodell für die Bundesrepublik Deutschland. VierteIJahrheft zur Wirtshaftsforschung 4 (1971), 268-277.

Benassy, J.P., Fouquet, D. and Malgrange, P., 1975, Estimation d'une function de production aux générations de capital. Annales de l'INSEE 19, (1975), 3-52.

Bliss, C.J., 1965, Problems of Fitting a Vintage Capital Model to U.K. Manufacturing Time-Series. Paper presented to the World Congress of the Econometric Society. Rome.

Brown, M., 1967, The Theory and Empirical Analysis of Production. NBER. New York and London.

David, P.A. and van de Klundert, Th., 1965, Biased Efficiency Growth and Capital-Labor Substitution in the U.S., 1899-1960. American Economic Review 3 (1965), 357-394.

Edgren, G., Faxén, K.-O. and Odhner, C.-E., 1970, Lönebildning och samhällsekonomi. Stockholm.

Ferguson, C.E., 1965, Time-Series Production Functions and Technological Progress in American Industry. Journal of Political Economy 2 (1965), 135-147.

Görzig, B., 1976, Results of a Vintage-Capital Model for the Federal Republic of Germany. Empirical Economics 1:3 (1976), 153-166.

Den Hartog, H. and Tjang, H.S., 1976, Investment, Wages, Prices and Demand for Labour. De Economist 124 (1976), 32-55.

Isard, P., 1973, Employment Impacts of Textile Imports and Investment: A Vintage Capital Model. American Economic Review 63 (1973), 402-416.

Johansen, L., 1959, Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis, Econometrica 27, (1959), 157-176.

Jungenfelt, K.G., 1966, Löneandelen och den ekonomiska utvecklingen. Industriens Utredningsinstitut, Stockholm.

Kennedy, C. and Thirlwall, A.P., 1973, Technical Progress. In Surveys of Applied Economics Vol. 1, 115-176. London and Basingstoke.

Krantz, O. and Nilsson, C.-A., 1975, Swedish National Product 1861--1970. Kristianstad.

Kuipers, S.K. and Bosch, H.F., 1976, An Alternative Estimation Procedure of a Clay-Clay Type of Vintage Model: The Case of the Netherlands 1959-1973. De Economist 124 (1976), 56-102.

Phelps, E.S., 1963, Substitution, Fixed Proportions, Growth and Distribution. International Economic Review 4 (1963), 265-268.

Sandee, J., 1976, A Putty-Clay Model for the Netherlands. Paper presented to the European meeting of the Econometric Society 1976.

Solow, R.M., 1960, Investment and Technological Progress. In K.J. Arrow et al., Mathematical Methods in the Social Sciences, 1959. Stanford (1960), 89-104.

Solow, R.M., 1962, Substitution and Fixed Proportions in the Theory of Capital. Review of Economic Studies 29 (1962), 207-218.

Sutton, J., 1976, The Movement of Factor Shares under Conditions of Rapid Economic Growth. In Artis and Nobay (Ed.) Essays in Economic Analysis. Cambridge 1976, 98-126.

De Vries, A.S.W., 1973/74, Een fixed coeffients vintage model voor Nederland, Maandschrift Economie 38 (1973/74), 391-412.

Wynn, R.F. and Holden, K., 1974, An Introduction to Applied Econometric Analysis. London and Basingstoke.

Åberg, Y., 1969, Produktion och produktivitet i Sverige 1861-1965. Industriens Utredningsinstitut, Stockholm.

A DYNAMIC MODEL OF RESEARCH AND DEVELOPMENT EXPENDITURE

M. Ishaq Nadiri
New York University

INTRODUCTION

The issue of integrating the demand for research and development expenditure of the firm with its demand for conventional inputs such as labor and physical capital has not received sufficient attention. The need for such an undertaking is clear: R & D, like expenditure on plant, equipment, and labor, is an input to the production process and, therefore, an integral part of the overall decision framework of the firm.

The primary purpose of this study is to investigate the determinants and consequences of an increase in research and development expenditures in the context of a disequilibrium dynamic model of a set of input demand functions. By means of this model the following issues are analyzed:

(a) The short-run effects of changes in output and relative prices on demand for innovative activities, measured by stock of R & D expenditure, employment, and capital stock;

(b) The spill-over effects of disequilibrium in any of these inputs on demand for the other inputs;

(c) The effects of research and development and plant and equipment expenditures on labor productivity in the short, intermediate, and long runs; and

(d) The responses of the inputs of firms of different asset sizes to changes in relative prices and output changes and the pattern of interactions among their inputs over time.

The plan of the study is as follows. The rationale of the disequilibrium approach to the analysis of input demands is described in Section I. In Section II, the estimating equations, the characteristics of the data, and some estimation problems are described. The structural estimates of the model using data for sixty-two firms for the period 1965 to 1972 are presented and discussed in Section III, Part A. In Part B of this section, the structural estimates of the model fitted to samples of firms classified by their asset size are presented. The stability of the model is also examined. In Section IV, the cross-sectional differences among firms in their demand for inputs are noted and the over-time differences among input demands are analyzed. The long-run output and price elasticities of employment, research and development, and capital stock are also discussed in this section. The summary and conclusions are stated in Section IV.

I. THE RATIONALE FOR A DYNAMIC DISEQUILIBRIUM MODEL

Existing cross-section and time-series models of the determinants of R & D behavior assume fixed stocks of capital and labor.[1] Also, no allowance is made in the employment and investment literature for the fact that a firm's R & D activities will affect its cost structure and thereby affect its demand for labor and capital.[2]

---

[1] See M.I. Kamien and N.L. Schwartz, "Market Structure and Innovations: A Survey". Journal of Economic Literature, 13:1 (March 1975), 1-37.

[2] Some examples of such studies are M. Baily, "Research and Development Costs and Returns: The U.S. Pharmaceutical Industry," Journal of Political Economy, 80:1 (January/February 1972), 70-85; H.G. Grabowski, "The Determinants of Industrial Research and Development: A Study of Chemical, Drug and Petroleum Industries," Journal of Political Economy, 76:2 (March/April 1968), 292-306; M.I. Kamien and N.L. Schwartz, "Risky R & D with Rivalry," Annals of Economic and Social Measurement, 3:1 (January 1974), 267-77 and "Market Structure and Innovations: A Survey," op.cit.; E. Mansfield, The Economics of Technical Change (New York: Norton 1968) and E. Mansfield, J. Rapaport, J. Schnee, S. Wagner and and M. Hamburger, Research and Innovation in the Modern Corporation (New York: Norton, 1971); and F.M. Scherer, "Firm Size, Market Structure, Opportunity, and Output of Patented Inventions," American Economic Review, 55:5 (December 1965), 1097-1125; Du Rietz, A., "Industriforskningens utveckling och avkastning" (Industrial Research and Development - Growth and Returns). IUI Stockholm. 1975

That is, decisions with respect to the conventional inputs will depend on when and how vigorously the firm engages in innovative activities. In turn, a firm's demand for research and development effort will be affected by the magnitudes and characteristics of its capital and labor. In this type of interactive process, all the inputs are essentially variable and are only differentiated from each other by the degree of their flexibility or adjustment over time.

The dynamic model described below permits interaction among these inputs over time. The main feature of the model is that disequilibrium in any of the inputs has a spill-over effect on demand for other inputs in the short run, while in the long run all excess demands disappear and the spill-over effects vanish.[1] However, in the very short run, as the firm attempts to adjust its stocks of inputs, it will increase the utilization of its existing stocks to meet current demand. As the stock adjusts, the utilization rates return to their optimum levels.

The Model of the Input Demand Functions

Assume that the firm minimizes costs subject to a Cobb-Douglas production function with three inputs: labor (L), capital stock (K), and stock of research and development activities (R). The input and output prices are assumed to be exogenously given. More formally, the general problem considered is to minimize costs:

$$C = wL + cK + rR \tag{1}$$

subject to the production function

$$Q = AL^{\alpha_1}K^{\alpha_2}R^{\alpha_3}U^{\alpha_4}e^{\lambda t}, \tag{2}$$

[1] I recognize that the dynamic input and output paths are jointly determined, contingent on future product price expectations. But their joint estimation requires a full market theory not yet available. Therefore, I set the limited goal of estimating optimum input paths consistent with an optimum and given output path. This allows me to concentrate on interactions among changes and on factor substitution.

where w, c, and r are, respectively, the user costs associated with employment, stock of plant and equipment, and stock of research and development. Q is the level of output, A is a constant, and α_1, α_2, and α_3 are the long-run output elasticities of the inputs; λ is the rate of disembodied technical change. We have assumed that the input utilization rates are functions of an overall rate of utilization, U. Also note that the utilization rate, U, does not enter the cost function explicitly, but implicitly through the rate of depreciation, δ, of capital stock. Depreciation depends on the rate of utilization, U, as well as time, i.e., $\delta = \delta(U,t)$.

The user costs are defined to include the purchase price, the opportunity costs of funds, depreciation expenses due to utilization and passage of time, tax considerations, and capital gains. For example, the user costs of capital goods can be stated as:

$$c = \frac{P_k(r + \delta)(1 - \bar{k} - vz + vzk')}{(1 - v)},$$

where P_k is the deflator for capital goods; r the cost of capital, measured as $r = i - (\dot{P}/P)^e$, where i is long-term interest rate and $(\dot{P}/P)^e$ is the expected change in prices; δ is the depreciation rate; P and $\dot{P}$ are the level and the absolute change in the general price level; $\bar{k}$ is the long tax credit amendment, k' is the effective rate of tax credit; z is the present value of depreciation, and v is the corporate tax rate. The user costs for labor services and for research and development efforts are in principle similar to c. The Lagrangian method for minimizing costs (1) subject to the production function (2) will yield the long-run solution of the determinants of the inputs.[1] That is,

---

[1] See M.I. Nadiri and S. Rosen, A Disequilibrium Model of Demand for Factors of Production (New York: National Bureau of Economic Research, 1973), pp. 19-21, for derivation of these expressions.

$$
\begin{aligned}
y_1^* &= L = g_1(x^*, \bar{P}) \\
y_2^* &= R = g_2(x^*, \bar{P}) \\
y_3^* &= K = g_3(x^*, \bar{P}) \\
y_4^* &= U = g_4(P)
\end{aligned} \qquad (3)
$$

where $\bar{P}$ is a vector of the relative prices of inputs, and the coefficient of x^* is $1/\rho = (\alpha_1 + \alpha_2 + \alpha_3)$, the reciprocal of returns to scale parameter. Assuming that the adjustment cost of each input is proportional to the gap between its long-run equilibrium and actual levels and is also affected by the disequilibrium of the other inputs, i, it can be shown that the approach to the long-run equilibrium of the system of inputs is approximated by the following set of differential equations:[1]

$$y_{it} - y_{it-1} = \sum_{j=1}^{4} \beta_{ij} \left[g_i(x_t, \bar{P}_t) - y_{jt} \right] + v_{it}, \quad (i=1,\dots, 4) \quad (4)$$

where β_{ij} is a non-diagonal matrix of adjustment coefficients and $v_1, \dots, v_4$ are random terms with zero means and variance-covariance matrix Ω. From the generalized adjustment model (4) we can find (a) the short-term impact of changes in output and relative input prices, (b) the transition or distributed lag patterns of the inputs to a change in these variables, and (c) the long-run price and output elasticities of the inputs.[2] Since the technical details of these problems are discussed elsewhere,[3] we may state that the short-term transitory responses are calculated by computing $[I-(I-B)]^{-1}$ and the long-run elasticities by computing $A[I-B]^{-1}$; $B = [\beta_{ij}]$ is the non-diagonal matrix of adjustment coefficients, Z is the lag operator, and A is the matrix of the coefficients of the exogenous variables.

---

[1] Ibid., p. 55.

[2] Ibid., pp. 24-39 for details.

[3] M.I. Nadiri and S. Rosen, "Interrelated Factor Demand," American Economic Review, Part 1, 59:4, (September 1969).

II. ESTIMATING EQUATIONS, DATA, AND ESTIMATION PROBLEMS

The model specified in Section I has been estimated using cross-section and time-series data on sixty-two firms for the period 1965-72. The main source of our firm data is the Compustat tapes. The sixty-two firms are drawn from five industries: five from Metal Extraction (SIC 10), twenty-eight from Chemicals and Allied Products (SIC 28), twelve from Non-Electrical Machinery (SIC 35), eight from Electrical Equipment and Supplies (SIC 36) and nine from Instruments (SIC 38). Thus, our sample is dominated by firms in the Chemical and Allied Products categories.

The empirical specification of the model differs somewhat from (4). The user costs of labor and research and development have been omitted due to lack of suitable data. The real wage rates for the appropriate two-digit industries are used as a proxy for these two user-cost variables. The user cost of capital for each firm is approximated by a measure constructed for the total manufacturing sector. The output prices are not available at the firm level; therefore, we have used wholesale price indices of the two-digit manufacturing industries as deflators for output, nominal wage rates, and the user cost of capital.

We may best view research and development in the context of the services of a given stock of R & D to the production of current output. Reliable estimates of the benchmark and depreciation rates for R & D at the individual firm level are not available. We constructed the stock of R & D by assuming an arbitrary depreciation rate of 10 % per annum for each firm. The 1965 R & D investment in constant dollars is used as the benchmark for those firms that did not report any figures prior to 1965, while for firms with more extensive data, the first year of consistent reporting was chosen as the benchmark.[1] Capital stock series for

---

[1] The expressions were also run with the flow measure of R & D expenditure. The overall results were generally similar to those reported in Table 1.

R & D and plant and equipment for each firm was constructed by the recursive formula,

$$K_{it} = I_{it} + (1-\delta_i)K_{it-1} \quad (i=1,\ldots, 62) \tag{5}$$

where I_{it} is the deflated individual firm expenditure on R & D or new plant and equipment; the deflator used for converting nominal expenditure series on R & D and plant and equipment into constant dollars is the deflator for plant and equipment (1958=100). δ_i are the individual firm depreciation rates calculated for plant or equipment as the ratio of depreciation expenses to the benchmark capital stock obtained from the firm's balance sheet. As noted earlier, the depreciation rates for R & D are assumed to be fixed at 10 %. The employment data refer to total employment of each firm. Unfortunately, it is not possible to break this aggregate series into production and non-production or scientists and engineers, etc. Similarly, it is not possible to separate research and development expenditures into privately and publicly financed categories.

The specific estimating equations used are,

$$\begin{aligned}
L_t &= \alpha_0 + \alpha_1 Q_t + \alpha_2 (w/c)_t + \alpha_3 L_{t-1} + \alpha_4 R_{t-1} + \alpha_5 K_{t-1} + \alpha_6 U_{t-1} + \varepsilon_1 \\
R_t &= \beta_0 + \beta_1 Q_t + \beta_2 (w/c)_t + \beta_3 L_{t-1} + \beta_4 R_{t-1} + \beta_5 K_{t-1} + \beta_6 U_{t-1} + \varepsilon_2 \\
K_t &= \gamma_0 + \gamma_1 Q_t + \gamma_2 (w/c)_t + \gamma_3 L_{t-1} + \gamma_4 R_{t-1} + \gamma_5 K_{t-1} + \gamma_6 U_{t-1} + \varepsilon_3 \\
U_t &= \delta_0 + \delta_1 Q_t + \delta_2 (w/c)_t + \delta_3 L_{t-1} + \delta_4 R_{t-1} + \delta_5 K_{t-1} + \delta_6 U_{t-1} + \varepsilon_4
\end{aligned} \tag{6}$$

where all the variables are in logarithms; R_t is the measure of research and development expenditures; L_t and K_t are the levels of employment and capital stock of the firm; Q_t is the level of output; (w_t/c_t) is the ratio of nominal wage rate to the user cost of capital goods. U_t is the rate of utilization of the appropriate two-digit industries used as a proxy for firms' utilization rate; R_{t-1}, L_{t-1}, and K_{t-1} are the lagged dependent variables; and ε_1, ε_2, ε_3 and ε_4 are the error terms.

The adjustment processes are embedded in the coefficients of the lagged dependent variables. The own-adjustment coefficient in each equation can be obtained from the regression coefficient associated with the lagged dependent variable and the cross-adjustment coefficients from the regression coefficients related to the lagged values of other dependent variables. For example, in the first equation of (6), the own-adjustment coefficient is $\hat{\beta}_{11} = (1-\hat{\alpha}_3)$ and the cross-adjustment effects of disequilibria in R & D and plant and equipment on employment are measured by $-\beta_{12} = \alpha_4$ and $-\beta_{13} = \alpha_5$. Then,

$$B = \begin{bmatrix} \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \gamma_3 & \gamma_4 & \gamma_5 & \gamma_6 \\ \delta_3 & \delta_4 & \delta_5 & \delta_6 \end{bmatrix}$$

constitutes the 4 x 4 non-diagonal adjustment matrix which traces the interdependence of the adjustment paths of the three inputs and the utilization rate over time.

Before estimating these equations, the problem of heteroskedasticity in our sample had to be considered. Except for the three aggregate industry-wide variables w_t, c_t and U_t, the remaining variables in (6) are specific to each firm. Error variance for large firms will substantially exceed those for small firms, and, therefore, there is the possibility that the cross-section, within-cell regression functions will have unequal error variances. As is well known, there are two ways to handle this possibility: the first is to test for the existence of heteroskedasticity among firms and eliminate the statistically significant outliers; the second is to transform variables so that the error variances will be homogeneous.[1] We have followed the second alternative, and now have two options: (1) a log transformation of the variables to equalize

---

[1] See E. Kuh, Capital Stock Growth: A Micro Econometric Approach (Amsterdam: North Holland Publishing Company, 1963), pp. 91-98.

the error variances on the assumption that they are strictly proportional to the size of the independent variables; or (2) fitting the model in the ratio form, which means dividing the firm-specific variables by an appropriate scale variable such as the total assets of the firm. Though we have used both of these procedures (using total deflated assets of the firms as the denominator in the ratio form of the model), we shall report only the logarithmic results.

Another important estimation problem that arises immediately is whether or not to impose the implicit constraint on the adjustment coefficients of model (6). If the adjustment coefficients are unconstrained, one of the two hypotheses about the production function is implied: (1) if the production function constraint always holds as an equality, then the adjustment process implies output to be endogenous during the adjustment period; and (2) on the other hand, if output is taken as exogenous, independent adjustments imply that firms may not be on their production functions. The values of the adjustment coefficients, then, will determine whether the firms are inside or outside of their production surface.

We have not imposed the necessary constraints on the adjustment coefficients mainly because the underlying data are unreliable. Instead, we have assumed that output is endogenous and have examined the unconstrained estimates of the adjustment coefficients to see whether the constraints implied by the model are met. The structural equations for each input are estimated by two-stage least-squares and the characteristic roots of matrix B are examined to check whether the implicit constraints are reasonably met.

III. THE STRUCTURAL ESTIMATES

The model is estimated using the variance components technique in pooling cross-section and time-series data developed by G.S. Maddala.[1]

[1] G.S. Maddala, "The Use of Variance Components Methods in Pooling Cross-Section and Time-Series Data", Econometrica, 39:2 (March 1971), 341-357.

This method allows estimating the cross-section and time-series effects separately and generates generalized least-squares estimates of the parameters of the model.[1]

The model (6) is estimated using the overall sample of sixty-two firms and three sub-samples: twenty-eight firms with total assets below \$300 million, twenty firms with assets greater than \$300 million but smaller than one billion dollars, and fourteen firms of over one billion dollars in total assets. Estimation of the model using the stratified samples should provide a test of its stability and insight into whether firms of different sizes differ in their input decisions. We have also estimated both the ratio and logarithmic forms of (6) for all four samples. Only the generalized least-squares estimates of the model in logarithmic form are presented here.

A. Structural Estimates for the Overall Sample

The results in Table 1 are the generalized least-square estimates with cross-section and time dummies. Note that $\hat{Q}_t$ is the estimated value of the output variable Q_t.

The results indicate a consistent picture: most of the coefficients were generally statistically insignificant in both the OLS and GLS versions, the results of the ratio and logarithmic forms of the model were fairly similar, and the signs of the coefficients of all but a few variables remained stable in the various versions of the model.

As can be seen from Table 1, the statistical goodness-of-fit of the model—measured by R^2, sum squares errors (SSR), and estimated variance of errors (EEV) — is very good. A separate test using

[1] The computer program based on this technique generates four regressions: the Ordinary Least Squares (OLS), Generalized Least Squares (GLS), which does not take account of cross-section and time effects, the Least Squares plus dummy variables (LSDV), which does take account of these effects, and finally, the Generalized Least Squares with dummy variables (GLSDV).

Table 1. Generalized Least Squares Estimates of the Model in Logarithmic Form
Period: 1965-72

Independent Variables	Generalized Least Squares Equations			
	Log L_t	Log R_t	Log K_t	Log U_t
C_0	-.3458 (-.5135)	.6793 (1.889)	-.3035 (-.8140)	-.2013 (-2.077)
Log $\hat{Q}_t$	.3355 (5.482)	.1970 (7.614)	.2279 (5.758)	.0290 (2.933)
Log $(w/c)_t$	-.1742 (1.6855)	-.2418 (1.876)	.0254 (.1773)	.0300 (.8151)
Log L_{t-1}	.5173 (8.507)	-.0904 (3.422)	-.0353 (.9482)	-.0253 (2.745)
Log R_{t-1}	.0997 (2.75)	.6999 (42.40)	-.0046 (.2094)	.0074 (1.34)
Log K_{t-1}	-.0544 (1.62)	.0804 (5.391)	.8175 (40.33)	-.0099 (1.95)
Log U_{t-1}	-.3859 (2.220)	-.1772 (2.52)	-.4388 (3.565)	.6504 (20.42)
R^2	.9283	.9767	.9878	.8851
SSR	.3469	.3531	.3475	.3353
DF	365	365	365	365
EEV	.0105	.0017	.0054	.00033

the TSP regression program indicates that the Durbin-Watson test values were about 2.0 for each of the equations. However, this test is not only biased when a lagged dependent variable is included as an explanatory variable but also may not be invariant with respect to the ordering of the firm data in our sample.

The estimates in Table 1 indicate the immediate responses of the inputs to changes in output, relative input prices, their own lagged values, and cross-adjustment effects of other inputs. The coefficient of output is positive and statistically significant in each equation. The output elasticities indicate that changes in output have the strongest effect on employment (.34), followed by stocks of capital goods and research and development. The output elasticity of the utilization rate, U, which should be very high, is rather small. The explanation for this is that our measure of the utilization rate is an industry measure which may not respond greatly to movements of demand of the individual firms. The relative price variable is also statistically significant and negative in both research and development and employment equations; it has the correct positive sign, but is not statistically significant in the utilization equation.

The own lag coefficients of the three stock variables indicate that employment adjusts very rapidly (1 - .52 = .48), followed by stock of research and development expenditures, (1 - .70 = .30), while capital stock adjusts very slowly (1 - .82 = .18). These patterns of adjustment are consistent with our a priori notion and previous results. They suggest, if we ignore the spill-over effects, an average lag of a year for employment, two-and-a-half years for research and development, and about four years for the capital stock.[1] The adjustment coefficient for the utilization rate is unexpectedly long. Again, part of the reason is that U is an industry measure and cannot be explained readily by movements of firm data. There are significant cross-adjustment effects in each

---

[1] These calculations are only very tentative for the adjustment patterns are interdependent and this interdependency cannot be ignored.

demand equation, though of varying magnitudes. These are calculated as $-\hat{\beta}_{ij}$, $i \neq j$; that is, the negative of the cross-adjustment coefficients shown in Table 1. For example, $-\hat{\beta}_{1j}$, $j = 2,3,4$ measures the effects of excess demand in employment on stocks of research and development and capital and on the utilization rate; they are shown by the coefficients in row L_{t-1} in Table 1. The signs and magnitudes of the cross-adjustment coefficients vary among the equations, indicating an asymmetrical and varying disequilibrium effect. As noted, the direction of these effects will be the opposite of the signs of the coefficients shown in Table 1.

(i) Excess demand for labor has a strong positive effect on the utilization rate and stock of research and development. It also affects demand for capital goods positively, but the effect is not statistically significant. Thus, excess demand for labor increases the utilization rate and demand for plant and equipment and R & D.

(ii) Excess demand in stocks of research and development has a strong negative effect on demand for labor; its impact on capital stock is positive but not very significant; its effect on the utilization rate though positive is barely significant statistically. Thus, disequilibrium in R & D capital reduces demand for labor but increases that of physical capital, implying a complementary relation with labor and a substitutional relation with physical capital.

(iii) Excess demand for physical capital has statistically significant positive effects on demand for labor and the utilization rate, while it has a strong negative and statistically significant impact on demand for research and development expenditures. These patterns of response suggest a short-run complementary relation between stocks of capital goods and research and development and a substitutional relation with employment.

(iv) The cross-effects of the rate of utilization on the demand for employment, research and development, and capital goods are all positive and statistically significant. That is, disequilibrium in the utilization rate leads to increased demand for productive inputs.

(v) These disequilibrium effects suggest that when the firm faces excess demand in one of its inputs it responds by increasing its rate of utilization and adjusting its demand for other inputs. Thus, strong feedbacks and dynamic relations exist among the inputs in the short run.

From these results, we conclude that there are strong and statistically significant short-term effects of changes in output and input prices on research and development, employment, and investment demands of the firm. Also, there are some lags in achieving the desired levels of these inputs. The lags are due not only to factors generating disequilibria in the specific input's own market but also to disequilibria in other inputs. Not only do dynamic feedback or spill-over effects among these three inputs exist, but they tend to be asymmetrical in character. The utilization rate serves as a buffer, allowing the firm to change its stocks of input. That is, when current demand increases, firms utilize their existing stocks of inputs more fully first and, then, if the demand is perceived to be fairly permanent, adjust their stocks of inputs.

B. Structural Estimates for the Sub-Samples

The results in Table 1 are essentially repeated when the model is fitted to the three sub-samples mentioned earlier. The structural estimates for the sub-samples are presented in Table 2. The striking over-all conclusion that emerges from a comparison of the results in Tables 1 and 2 is the stability of the model in terms of signs and significance of the coefficients, and the goodness-of-fit statistics, such as R^2 and sum-of-squares errors. The magnitudes and statistical significance of the coefficients vary somewhat across different asset sizes. The output variable is statistically significant in all of the regressions; the magnitudes of the coefficients are larger and similar to that of the overall sample (except for the employment equations) for firms with assets greater than one billion and those with assets less than 300 million dollars. For the medium-size firms, the short-term responses of the inputs to changes in output is somewhat smaller, except in the employment

Table 2. Generalized Least Squares Estimates of the Model in Logarithmic Form for Three Samples of Firms Period 1965-72

In-dependent variables	Twenty-Eight Small Firms Equations				Twenty Medium Size Firms Equations				Fourteen Large Firms Equations			
	Log L_t	Log R_t	Log K_t	Log U_t	Log L_t	Log R_t	Log K_t	Log U_t	Log L_t	Log R_t	Log K_t	Log U_t
C_0	-.0175	.8677	-.4155	-.1795	-3.4175	.1419	.3517	-.3415	.6690	.0917	.7089	-.0591
	(-.0198)	(1.888)	(-.7361)	(-1.518)	(-2.031)	(.4093)	(.5351)	(-1.432)	(.4639)	(.1921)	(1.059)	(-.1638)
Log $\hat{Q}_t$	.1828	.2093	.2459	.0231	.5640	.1072	.1977	.0341	.9619	.0913	.2290	.0428
	(2.003)	(5.755)	(3.408)	(1.725)	(5.274)	(2.424)	(3.821)	(1.745)	(5.692)	(1.9345)	(3.7395)	(1.1480)
Log $(w/c)_t$	-.1410	-.3216	.0615	.0319	.6427	-.0057	-.0190	.0797	-.6367	.0718	-.0990	.0024
	(-.4254)	(-1.975)	(.2779)	(.6954)	(1.187)	(-.0396)	(-.0821)	(.9506)	(-1.285)	(.4457)	(-.4410)	(.0185)
Log L_{t-1}	.6587	-.1394	-.0324	-.0175	.1329	.0038	-.0342	-.0331	.1581	-.0241	-.0589	-.0333
	(7.309)	(-3.7)	(-.4831)	(-1.3793)	(1.015)	(.0832)	(-.5525)	(-1.571)	(1.417)	(-.7772)	(-1.4703)	(-1.3625)
Log R_t	.1028	.6569	-.0164	.0042	.4347	.9365	.0768	.0157	-.0386	.7835	-.0101	.0141
	(2.1625)	(32.981)	(-.472)	(.6090)	(3.7784)	(35.532)	(1.605)	(1.1005)	(-.4965)	(34.787)	(-.3428)	(.7913)
Log K_{t-1}	-.0649	.1079	.7994	-.0114	-.2485	-.0801	.6827	-.0185	-.3430	.0596	.7417	-.0350
	(1.516)	(6.2145)	(24.609)	(1.8116)	(-1.2013)	(-2.831)	(15.182)	(-1.1888)	(-3.0043)	(1.6578)	(15.172)	(-1.4667)
Log U_{t-1}	-.3049	-.1769	-.5847	.6607	-.3088	-.1775	-.1429	.6272	-1.3089	-.2421	-.4469	.6265
	(-1.0693)	(-1.683)	(-2.466)	(13.85)	(-1.2013)	(-1.481)	(-1.1042)	(10.988)	(-3.922)	(2.534)	(-3.546)	(8.3232)
R^2	.9648	.9923	.9897	.9443	.9632	.9968	.9835	.9668	.7802	.9798	.9515	.7675
SSR	.1562	.1608	.1568	.1537	.1099	.1062	.1079	.1037	.07604	.07927	.07672	.07610
DF	161	161	161	161	113	113	113	113	77	77	77	77
EEV	.0134	.0018	.0105	.0003	.0052	.0014	.0013	.0002	.0078	.0006	.0009	.0004

equation. The relative price variable (w/c) has the correct sign in most cases, but in most of the regressions its magnitude and statistical significance vary. However — again, with the exception of the employment equations — the coefficients of the relative price variable are statistically insignificant.

The own and cross-adjustment coefficients are quite strong in some of the regression equations in Table 2. The asymmetrical pattern noted for the whole sample holds in the sub-sample regressions as well; the magnitudes of the own and cross-adjustment coefficients, however, vary among firms with different asset sizes. The weakest links in the feedbacks among the input disequilibria are observed in the effects of excess demand for R & D of firms with assets over one billion dollars. Disequilibrium in capital stock has strong effects on the demand for research and development of firms in all asset categories. The utilization rate affects positively the demand for all the inputs, as we noted earlier, for the whole sample of firms. The employment disequilibrium has a fairly weak effect on demand for R & D and capital stock in the medium and large-size firms.

To test the stability of the model across the asset classifications, we computed the relevant F statistics for each set of input demand equations:

$$F = \frac{SSE_T - (SSE_{14} + SSE_{20} + SSE_{28})/k}{(SSE_{14} + SSE_{20} + SSE_{28})/N-3k},$$

where SSR_T is the sum-of-squares errors from the regression for the 62 firms and SSE_{14}, SSE_{20}, SSE_{28} are the sum-square errors from the regressions for the sub-samples of firms. N is the overall number of observations and k is the number of the parameters estimated. The calculated F statistics for L, R, and K equations are 0.689, 0.9504, 0.8927, and 0.2652, respectively, and the critical value of F (7,344) at one percent level of confidence is 2.69. Therefore, the null hypothesis of an unchanging structure of demand functions for labor, research and development, and capital goods cannot be rejected.

The Cross-Section and Over-Time Differences among Firms

The analysis of variance employed in estimating the demand equations permits testing whether cross-section and time-series differences exist among our sample of firms in their input decisions. We have calculated the F statistics based on the estimates generated by the least squares plus dummy variables (LSDV) of the analysis of variance. The results in Table 3 pertain to logarithmic form of the model using the entire sample and three sub-samples of firms. They indicate an interesting pattern: Substantial cross-sectional differences exist among firms with respect to all of the inputs and, except for the demand for research and development expenditure in the small and medium-size firms, all the input functions also vary over the span of time considered.

It is difficult to state precisely the causes of the cross-section and time-series differences among the samples of firms in their input decisions. The cross-sectional difference may arise due to the differences in the characteristics of firms, such as being in different industries, producing different types of products, having different degrees of monopoly or monopsony in the markets, etc. The over-time differences may be due to differing adjustment processes, responses to external stocks, and technological changes. Though very desirable, a closer look into the sources of these differences in input demand functions of the firms is beyond the scope of our present research.

IV. CONCLUSIONS AND SUMMARY

The results presented in this paper indicate that the firm's employment, capital accumulation, and research and development decisions are closely intertwined and that a dynamic interaction process seems to underlie these decisions. The research and development activities of the firm, like its demand for labor and capital, are influenced significantly by changes in output and relative input prices. The long-run output elasticities of the inputs, especially those of labor and research and development, are quite similar and suggest a slight increasing return to

Table 3. Values of F-Statistics from Analysis of Variance for the Entire Sample and Three Subsamples of Firms for the Period 1965-1972[a]

Group	Dependent variable	Effects: Cross section	Time-Series
Overall sample: Sixty-two firms	L_t	63.5776	5.2096
	R_t	245.087	10.3005
	K_t	21.6712	7.6988
	U_t	19.5848	173.861
Fourteen large firms	L_t	81.5190	3.5241
	R_t	161.635	9.2359
	K_t	204.780	18.0309
	U_t	45.4924	31.5152
Twenty medium-size firms	L_t	481.675	62.9452
	R_t	38.3298	1.1649
	K_t	267.811	9.7051
	U_t	47.5290	90.3784
Twenty-eight small firms	L_t	357.410	41.1140
	R_t	321.932	3.6381
	K_t	14.5643	2.5744
	U_t	16.1273	90.8643

[a] The critical values of F for the cross-section estimates at .05 are approximately: $F(61,305) = 1.47$ for the entire sample; $F(13,65) = 2.42$ for the fourteen large firms; $F(19,95) = 2.09$ for the medium-size firms and $F(27,135) = 1.85$ for the twenty-eight small firms. The critical values of F for the time-series estimates at .05 are, respectively, $F(5,305) = 3.09$, $F(5,65) = 3.29$, $F(5,95) = 3.20$, and $F(5,135) = 3.17$.

scale in production. Both labor productivity and investment demand of the firms are affected significantly by their research and development expenditures. These results are in contrast to the findings of the familiar investment and employment functions which often have ignored the explicit role of research and development. We find that the demands for the three inputs are stable when firms are stratified by asset size; however, there is evidence of cross-sectional and over-time differences among firms in their input decisions. The causes of such differences are not explored at the present.

To improve our empirical results, some of the shortcomings of our present data base have to be remedied. It would be useful to enlarge our sample of firms both in numbers and in distribution over industry classifications. The data for wage rates and user costs of capital could be improved by obtaining more disaggregate measures of these variables; there is a need to construct the rental price of research and development activities and to develop better capital stock measures for R & D at the firm level. It would be useful, if data permit, to classify the firms by industry and contrast the inter-industry differences in employment, capital accumulation, and research and development expenditures. A test could also be developed to estimate the sensitivity of firms' demand for inputs to changes in aggregate economy variables and to examine more closely the cyclical characteristics of these input demand functions.

REFERENCES

Baily, M.N., 1972, "Research and Development Costs and Returns: The U.S. Pharmaceutical Industry", Journal of Political Economy. January/February 1972. 80:1:70-85.

Bischoff, W., 1971, Brookings Papers on Economic Activity.

Brainard, W. and Tobin, J., 1968, "Pitfalls in Financial Model Building", American Economic Review. May 1968, 58:99-122.

Branch, B., 1973, "Research and Development and its Relation to Sales Growth", Journal of Economic Business. Winter 1973, 25:2:107-11.

Du Rietz, A., 1975, Industriforskningens utveckling och avkastning (Industrial Research and Development - Growth and Returns). IUI. Stockholm

Gordon, J., 1971, "Inflation in Recession and Recovery", Brookings Paper on Economic Activity, 1971:1, Table A-1, p. 148.

Grabowski, H.G., 1968, "The Determinants of Industrial Research and Development: A Study of Chemical, Drug and Petroleum Industries", Journal of Political Economy, March/April 1968, 76:2:292-306.

Kamien, M.I. and Schwartz, N.L., 1971, "Expenditure Patterns for Risley R & D Projects", Journal of Applied Problems, March 1971, 8:60-73.

\- , 1974, "Risky R & D with Rivalry", Annals of Economic and Social Measurement, January 1974, 3:1:267-77.

\- , 1975, "Market Structure and Innovations: A Survey", Journal of Economic Literature, March 1975, 13:1:1-37.

Kuh, E., 1963, Capital Stock Growth: A Micro Econometric Approach, Amsterdam: North Holland Publishing Company.

Maddala, G.S., 1971, "The Use of Variance Components Methods in Pooling, Cross Section and Time Series Data", Econometrica, March 1971, 39:2:341-57.

Mansfield, E., 1968, The Economics of Technical Change, New York: Norton.

, Rapaport, J., Schnee, J., Wagner, S. and Hamburger, M., 1971, Research and Innovation in the Modern Corporation. New York: Norton.

Nadiri, M.I. and Rosen, S., 1969, "Interrelated Factor Demand Functions", American Economic Review, September 1969, 26:4:457-71

— , 1972, "An Alternative Model of Business Investment Spending". Brookings Paper on Economic Activity, 1972:3, 576

— , 1973, A Disequilibrium Model of Demand for Factors of Production. National Bureau of Economic Research. New York

Nerlove, M., 1967, "Notes on the Production and Derived Demand Relations Included in Macro-Econometric Models," International Economic Review (June 1967), 8:223-42

Scherer, F.M., 1965, "Firm Size, Market Structure, Opportunity, and Output of Patented Inventions," American Economic Review (December 1965)

Sims, C.A., 1974, "Output and Labor Input in Manufacturing," Brookings Papers on Economic Activity, 3:695-736

Standard & Poors Corporation, Compustat Tapes, various issues

U.S. Department of Commerce, Survey of Current Business, various issues

U.S. Department of Labor, Earnings and Employment for the U.S., 1909-65. Washington, D.C. 1965

— Monthly Labor Bulletin, various issues

RELATIVE PRICE CHANGE AND INDUSTRIAL STRUCTURE - THE "NORWEGIAN CASE"

Gunnar Eliasson
IUI, Stockholm

1. INTRODUCTION

Exogenously induced growth stimuli to an economy are not always 100 percent good things if considered within a sufficiently long time horizon. If these stimuli are too strong and/or too sudden, the economy gets overheated and price mechanisms become disorderly. The information content of price signals changes character when interpreted by old (decision) rules of thumb. Important decisions can go wrong at the production level and in the pricing of factors of production, but most importantly on the investment side. Investment takes a long time to be decided on, and takes a long time to affect the economy, and mistaken decisions take an equally long time to be corrected.

This paper was originally conceived as an illustration of what happens to information handling and decision making in the market mashinery of an advanced industrialized economy like Sweden when subjected to a double experience of the Norwegian type; the North Sea oil discovery in conjunction with a later, sudden and very strong, maintained price increase in that same sector.

The North Sea oil discovery -- a tremendous growth impulse -- has also aroused public concern in Norway about the indirect effects of relative price changes (and the consequent wage drift) on other sectors. We will simulate a particular and stronger version of the Norwegian experience on the micro-to-macro model of the Institute loaded with data from a Swedish like economy. The elaborate treatment of the supply side in the short and long runs for each firm that makes total economic growth fully endogenous within an upper technology constraint makes this model particularly useful for the analysis of this kind of problem.

The whole raw material sector of this model version of Sweden will be subjected to both a price and a "technological" shock experience of a kind similar in principle to what the Norwegian economy has been subjected to. This is the reason why we have given the paper the subtitle: The Norwegian Case, even though the numerical data as such do not pertain to the Norwegian economy. Even with this explanation, the title may still be considered somewhat misleading. While the disturbing influence on wage setting was at first expected to originate directly in the fast-expanding oil producing sector,[1] it is now more commonly seen as emanating from an excessively expanding public sector that feeds on the "tax" proceeds from the oil sector. The principal results are, however, the same whichever viewpoint one adopts.

---

[1] See e.g. chapters 8 and 9 in Parliamentary Report No.25 (Petroleum Industry in Norwegian Industry), Ministry of Finance,1973-74 and also Norsk industriutveckling och framtid, Norges Industriförbund, debatt- og studieheften, 1975, nr 8.

2. THE MODEL

The model can be most simply presented as a set of individual firm models aggregated to the national accounts level through an explicit labor and product market process, where all prices are endogenously determined, the whole system being encased in a Leontief-Keynesian macro framework. The total model integrates (micro) market theory and income determination theory in an unelaborate but effective way. The theory of the firm upon which the firm model is based was previously developed in Eliasson [1976a], and a fairly complete description of the model is found in Eliasson [1976b].[1] The model is now loaded with numbers to make it represent a Swedish like economy.

It is capable of simulating post war inflation patterns and growth trends for a spectrum of macro variables quite well. The cycles are, however, not well reproduced, and as this is being written (July 1977) a large data base job and much calibration work lie ahead. This means that we will restrict our comments and tabular material to periods not shorter than 5 years and the results to be reported on should be viewed as a numerical analysis of the theoretical properties of a model economy similar to the Swedish economy.

The most important exogenous variables of the Swedish micro-to-macro model are a) the rate of change in labor productivity of new equipment, b) foreign prices (one index for each of four markets), and c) the nominal rate of interest. The rate of industrial growth is therefore endogenous through an endogenous investment function with each firm. Growth is bounded above by the extent of investment and by the new (exogenous) technology brought in by new investment.

---

[1] A very compact presentation of the model can also be found in Eliasson [1977]. A report on the new, extended version, to be described below, was under preparation when this paper was read and has now been published in Eliasson (ed.) [1978a].

Investment depends heavily on business profits which in turn depend importantly on how correctly firms interpret current price, wage and profit signals and transform these into expectations. Profit targets of individual firms are set on the basis of past experience. If performance is improving, targets are gradually raised and conversely if performance is declining. Zero production is the lower bound of the activity level.

Total demand is completely endogenized. Wages, as determined in the labor market, feed back through a Friedman (Permanent income)-Stone type expenditure system. Household saving is treated as one expenditure category and durables are entered through a stock demand device.

Export supplies from the Swedish production system respond to relative foreign-domestic price differentials and similarly on the import demand side.

In fact, all business decisions at the firm level are in terms of reactions to expected relative price movements or differentials that are checked against internal profit targets in the firm. This, in combination with the explicit "tâtonnement" process in the labor and product markets, the feedback of total income into demand and the dependence of investment on profit rate gives the total model economy several uniquely dynamic properties. Some of them will be investigated in this paper.

Three properties of the total macro system should be mentioned. First, we met with initial difficulties in finding a parameter specification that generates a growth development similar to experience in Sweden over the post-war period. When fed with the post-war exogenous input trends in foreign prices and productivity growth in new investment vintages we have now managed to make the model reproduce the post-war, long run growth trend in a chosen set of key macro variables such as industrial production, wholesale prices and profit margins. A general property of the system is, however, that this successful growth performance is built upon a quite delicate balance of factors. It is easily disturbed and then results in a downward bend in growth rates. If the simulation is allowed to go on further, growth gradually tapers off.

Second, such long term bends occur even though the underlying exogenous upgrading of technology continues steadily on the same growth trend. We have in fact been able to generate very diverging long run 20-50 year growth trends on the same assumptions as to technical change (in MTEC in (6), (7) and (8) below), using different market performance and cyclical assumptions.

Third, no irregularities occur if exogenous inputs stay within the normal range of variation. However, if the model is subjected to shocks ("positive" or "negative") a strong macro response of expected type follows, but after some more years macro activity levels off inevitably and occasionally falls drastically. This reversal effect à la the Le Chatelier-Braun Principle of thermodynamic systems is everywhere present in the model. One could also say that the model responds with a typical business cycle to exogenously induced shocks. In cases when very strong reversal effects tend to develop, and where we have allowed the simulation to run long enough, activity levels eventually stabilize for a long time on a new, "normal" growth path below the one recorded in the reference run without the "shocks". This is so whether the original shock involved a positive or negative demand stimulus. We have come upon several instances in which a strong positive economic policy stimulant has worse long term effects than a more moderate "negative" policy measure. It all depends on the economic situation when measures are enacted and how they affect (disturb) the reliability of market price signals and the market allocation mechanisms. This is an interesting "asymmetric" property of the model. To my knowledge there is no systematic evidence available to shed light on the question of whether this is an empirically relevant property or not, except ad hoc observations about historical economic shock experiences, of which the present so called "oil crisis" is one.[1]

[1] Cf my paper "How does inflation affect growth?" in Eliasson (ed.) [1978a].

To understand the experiments to be reported on in this paper some features of the firm model have to be explained in some detail: These are a) the concept of international competitiveness used, b) the export and import functions, c) the expectations-profit targeting system and d) the production system.

a) International competitiveness

Competitiveness in the business world is invariably linked semantically to profits. To most people, however, international competitiveness of an economy as a concept would have a welfare implication in terms of the real income growth capabilities of the economy compared to other economies. If the degree of international competitiveness is defined from the welfare side as the capability of an economy to maintain a growth rate above some other country or group of countries the two concepts can be strongly linked together.[1] Ex post competitiveness is measured as an above-normal rate of growth for the country as a whole, and this is often the way the "phenomenon" as such is first observed. The next, natural step is to identify the determinants of this particular growth performance. The key indicators of supreme competitiveness normally listed are costs relative to the rest of the world, technical change, productivity change, etc. All come together by definition as elements in a relative profitability measure, and conventional opinion seems to be that there is a strong and monotonic relationship between profitability and economic growth. This essay will demonstrate that this is not necessarily and evidently true, except in a trivial ex post accounting sense.

b) Export and import determinants

In the model competition from abroad enters through the exogenous world market price level of each sector. Firms in the model (read: country of inquiry) face this price spectrum in domestic and export markets and are successful if they have a product mix

[1] As suggested in Eliasson [1972] pp. 129-133.

and a production structure that gives them a sufficient productivity performance (at the going wage etc., cost levels) to meet set profitability standards. In terms of the above argument the country is successful, or competitive, if these standards are such that a relatively high, sustained economic growth rate can be maintained.

Pratten [1976] has empirically illustrated the non-triviality of this statement. He finds that the Swedish economy has grown faster than the U.K. economy for a long time and that Swedish firms have exhibited substantially higher productivity measures than "matched" U.K. firms. Nevertheless, U.K. exhibits higher rates of return to capital.

Total market behavior in the entire model economy determines all domestic prices, including wages that go into the income and cost accounts of individual firms. Costs are, however, influenced by current productivity which is in turn (for each firm) partly influenced by the exogenously given rate of change in labor productivity (at normal capacity utilization levels) of new vintages of investment. Given this and its rate of return requirements (see below), each firm can calculate an output level that is compatible with profit targets at expected wages. All supply decisions together determine all prices and aggregate income (that enters as an argument of total private demand) and profits (that determine investment and capacity growth, see below), and so we have formed a dynamic link between all of the relevant determinants of profitability and economic growth. By doing so we can analyze the traditional indicators of international competitiveness and see to what extent there is the implied correspondence between their relative movements over time and the welfare indicators, like economic growth, that we are ultimately interested in. We are able, for instance, (in the Swedish micro-to-macro model) to study the somewhat surprising implications of a sudden price or technological upheaval in a large sector of an economy on the degree of competitiveness of individual firms as well as the material welfare of the entire economy. The link comes by way of the direct and indirect effects on all sectors of the economy of a windfall increase in the level of technology and the purchasing power of one sector, that allows the whole economy to draw on resources in foreign markets (at least for a while) on the basis of a temporary "land rent" or a transitory monopoly position.

Export functions (X) relate to each individual firm. Import functions (IMP) relate to markets. They are all expressed as ratios of total sales (exports) or total supplies (imports). Relative foreign-domestic price differentials the quarter before are the sole determinants of changes in these ratios.

$$\Delta X = f_1\left\{\frac{P(FOR) - P(DOM)}{P(FOR)}\right\},\ f_1' > 0,\ f_1'' < 0 \quad f_1(0) > 0 \qquad (1)$$

$$\Delta IMP = f_2\left\{\frac{P(DOM) - P(FOR)}{P(FOR)}\right\},\ f_2' > 0,\ f_2'' < 0 \quad f_2(0) > 0 \qquad (2)$$

FOR indicates foreign (exogenous)
DOM indicates domestic (endogenous in system)

The rationale for this simple formulation with no foreign demand factors is that the true decision variable relating to the question of where to sell must be relative profitability. For Swedish-based firms there is no reason to expect product costs to differ significantly between domestic sales and export sales when measured at the border passage. Hence product prices alone enter the decisions. With a long time series of short period (months or quarters) of price, X and IMP data it should be possible to estimate export and import price elasticities in a proper way. When observations refer to longer periods (say years) some of the volume responses to the price changes take place within the measurement period, making it difficult to quantify the importance of relative price changes properly. One obtains a better fit by including foreign demand variables like GNP or industrial production, although price and demand variables are not really compatible in the same formulation.

There is a strong self-regulatory feedback on the entire model economy from the export and import functions that also tends to keep the foreign accounts in balance in the longer term. The larger the gap between foreign and domestic prices the larger the share of domestic output that leaves the country, reducing (to begin with) domestic supplies and forcing up domestic prices to check the outflow. There is a mirror, supporting mechanism on the import side, and the whole process of course works in the other direction if we change the sign of the price gap.

c) Expectations and profit targeting

While foreign trade functions determine how world markets impact the outer surface of the economic system under study, expectations and profit targeting determine how the system responds internally. There is no use introducing formal specifications to explain in this brief context, since it would only detract attention from systems behavior as a whole, which is what matters. For a detailed understanding the reader is referred to Eliasson [1978a] chapter 4.

We will indicate only the main principles involved. Expectations functions of the feedback, error correction type refer to prices, wages and sales. Expectations determine ex ante calculations of profitability that guide the search for a production plan within the production system to be described below.

The profit targeting device is the criterion that indicates when a satisfactory plan has been obtained. Our formulation of the targeting device includes the conventional profit-maximizing device as a special case and hence is a more general criterion. It also has a better empirical foundation (see Eliasson [1976a]). Firms determine (on the basis of their own profit history)[1] what constitutes a feasible profit performance to use as a target. The target variable is the profit margin (frequently used within firms), and this corresponds to a long-run real rate of return

---

[1] and also by external information, say, by looking at the best performer in the market.

requirement (Eliasson [1978a] pp.58-69). It is complemented by various checks that prevent the firm from implementing this long term requirement too drastically in the short term. Targets can always be set higher and higher under the constraint that expected profits do not decrease, to approximate profit maximization. Since the nominal rate of return-interest rate differential determines the rate of borrowing and since total cash flows move investment spending as long as capacity is insufficient, it is easy to see how disorderly price signals in markets disturb firms' information system through their expectations functions. Erroneous decisions lead to a worsened profit performance to the detriment of growth.

d) The production system

The production system is essential for the supply properties of the entire model. Each period, each firm has its own transitory production frontier that determines the relationship between effective labor input and output. How these functions are estimated is described in Albrecht's paper in this conference volume. The production function is bounded above and marginal labor productivity is monotonically decreasing. It has the following mathematical form (somewhat simplified):[1]

$$Q = QTOP\cdot(1-e^{-\gamma L}) \tag{3}$$

QTOP is the horizontal asymptote towards which Q moves for unlimited increases in labor input (L). γ determines the bending of the curve (see below). Zero labor input means zero output.

The firm is currently operating on this production frontier or (mostly[2]) somewhere underneath it. If the current operating position does not satisfy profit margin targets at expected prices

---

[1] See further Eliasson [1978a] pp. 63-68.

[2] These are our results from the planning surveys of the Federation of Swedish Industries that supply the data needed to estimate the frontier and to position the firm underneath it. See further Albrecht's paper in this volume.

and wages the firm edges itself towards an improved (more productive) position closer to the frontier to the extent this is possible and as long as it does not diminish expected profits.[1]

The production frontier Q = f(L) is a soft surface in the sense that if the profit situation deteriorates enough, firms are capable of "doing better than normal" by a slack activating device (see Eliasson [1978a] pp. 13 and 71).

The reader should note that neither a capital stock nor a flow of capital services enter the momentary production frontier above. This production factor enters through the coefficients of the (Q,L) relationship, and these are supplied at startup time for a model simulation from individual firm data (available from 1975 from the planning survey of the Federation of Swedish Industries) and are updated by investment each period.

This updating takes place in the following manner.
Each period the (Q,L) frontier pivots down around the origin because of a lowering of QTOP due to economic wear and tear of equipment.

$$QTOP(t) = QTOP(t-1)*(1-\rho). \qquad (4)$$

The rate of depreciation (ρ) is exogenous.

Second, new investment both pivots (Q,L) in the opposite direction and bends it, due to improved technical qualities of equipment, through the following four equations:

$$\Delta QTOP = \frac{INVESTMENT}{P(DUR)} *INVEFF \qquad (5)$$

$$TEC(t) = \frac{QTOP(t-1) + \Delta QTOP(t-1)}{\frac{QTOP(t-1)}{TEC(t-1)} + \frac{\Delta QTOP(t-1)}{MTEC(t-1)}} \qquad (6)$$

$$\Upsilon(in(3)) = \frac{TEC}{QTOP} \qquad (7)$$

$$\frac{\Delta MTEC}{MTEC} = \text{Exogenous} \qquad (8)$$

---

[1] This search is quite complex. It is described in full detail for an earlier version of the model in Eliasson [1978a] and (will be) in full detail for this and a more sophisticated version of the total model system in a report currently being prepared.

INVESTMENT is expressed in current prices and allocated to the period when investment becomes operational. To handle this we currently use a third-order exponential delay function. P(DUR) is the appropriate deflator, endogenously determined in the model. INVEFF is a coefficient that determines the potential output (QTOP) yield from a unit of investment. It can be said to represent the marginal capacity-capital ratio. As such it should incorporate some exogenous information as to the qualitative upgrading of investment goods from a capital (not labor) augmenting point of view. For the time being we have not finally decided how to handle the amorphous concept of capital productivity in the model and have settled for a provisional and empirically reasonable approximation. In each quarter we approximate the new marginal output-capital ratio (= INVEFF) with the average ratio of value added in current prices to production equipment measured properly on a current replacement cost basis in the balance sheet. At each point in time this can be thought of as a conventional "technical coefficient". Both the numerator and the denominator are, however, updated in the model as to <u>volume</u> as well as <u>valuation</u> (price) by the events affecting the <u>firm</u> in the model. <u>This</u> means that a different development of product (i.e., the firm's) and investment goods prices affects INVEFF. It is not clear whether this is a desirable property or not. It is partly a technical price index problem.[1] The valuation principle choosen also mimics the way firms think about it in their internal accounting routines. This is important in this model context where measurements stretch all the way down to the production units. The major problem is, however, the approximation of the marginal ratio, with an estimate of the corresponding average ratio. In the future, however, the whole string of problems associated with this provisional approximation should go away, since we plan to estimate INVEFF directly using outside information.[2]

The harmonic average (6) above tells how the average technological position of the firm (TEC) is updated through a new vintage of investment.

The production function hence is of a putty-clay type with no explicit, aggregate capital stock measure. In diagrammatical terms we could say that a new (Q,L) relationship (3) of superior technical quality (MTEC > TEC and correspondingly a new γ) is superimposed on the old relationship, merged and stirred well to produce a new updated (Q,L) relationship. This means that

---

1 An analogous problem is faced when using time series of production volume and capital stock volume data to estimate capital output ratios. If the base year of the two deflators is changed the volume ratios are also changed.

2 The 1977 Planning Survey of the Federation of Swedish Industry collects an estimate from firms on INVEFF for 1977.

we do not keep each vintage of investment separate in the model.[1]

We have modelled the production system as it normally appears in firm planning and costing systems from which our measurements come, so this is the way we want to have it. The most frequent method in numerical planning in business firms is to bypass the problem of entering an explicit capital stock measurement by working with exogenously updated coefficients taken from the cost accounts (Eliasson [1976a] pp.296-300). The reason is of course the doubtful operational content of capital measures. Those who so desire can envision a shadow production function with aggregate capital stock (K) explicit. In this (Q,L,K) relationship the marginal product of labor approaches zero, and output is everywhere bounded above for unlimited labor inputs, which is a desirable property. In the explicit model of the firm, and as well in total industry, capital equipment enforces an indirect uppertime bound on output because investment goods are endogenously produced by the system.[2] This brings the upper bound back altogether upon labor input in the production process and the efficiency with which all resources are allocated by markets in the entire model economy. Zero labor input means zero output.[3]

To derive the shadow production function from equation (3) to (8) above we obtain a pair of partial differential equations that we have not been able to solve. Their properties can, however, be illustrated through numerical experimentation on the model. We have noted as a curiosity that whenever the model generates a smooth, horizontal trend in the profit share in output, Cobb-Douglas production functions always fit the synthetic time series data well. Not so if there is a sufficiently strong non-horizontal trend and/or if there are large deviations from a horizontal trend.

---

[1] The reason is of course the rapidly declining returns to cumbersome specification. Se further Albrecht's paper in this conference volume. When this paper is being finally edited (June 1978) we are working on a more sophisticated specification that will make it possible to approximate the vintage structure under steady state growth assumptions and also to make economic depreciations endogenous, much along the lines suggested by Bentzel in his paper in this conference volume.

[2] Also cf. Färe's paper in this conference volume.

[3] In this sense we have taken out the property of (for instance) the CES function that makes it possible to compensate one factor for the other when the elasticity of substitution is larger than 1 to the extent that output is then not bounded when labor is increased indefinitely, ceteris paribus. See e.g. Ferguson [1975] p. 103.

3. PROBLEM SPECIFICATION

To study the consequences of relative price changes on industrial structure we have performed the following experiments on the micro-to-macro model. We have subjected the raw material sector (14 percent of value added in total manufacturing 1975) to a sudden 40 percent exogenous (foreign) price increase. The relative foreign price change so obtained is maintained through a 20 year simulation, and constitutes the only difference in specification from the reference case.

This is a rather dramatic experience (albeith of a "positive" nature) for such a large sector.

We have repeated the same experiment in a softer mode, namely a 10 percent price change.

These examples have been chosen to illustrate the effect of subjecting an important export sector to a sudden price-induced increase in foreign demand like the oil price hike for oil producing sectors of an economy. We have also wanted to reproduce the case of a sudden discovery of oil. This is technically engineered by a sudden increase in potential output in the raw material sector, also this time by 10 percent. In all three cases the induced change happens in the second year. This is what happens to the model economy.

4. RESULTS FROM EXPERIMENTS - LONG RUN DEVELOPMENT OF MACRO ACTIVITY LEVELS (DESCRIPTION)

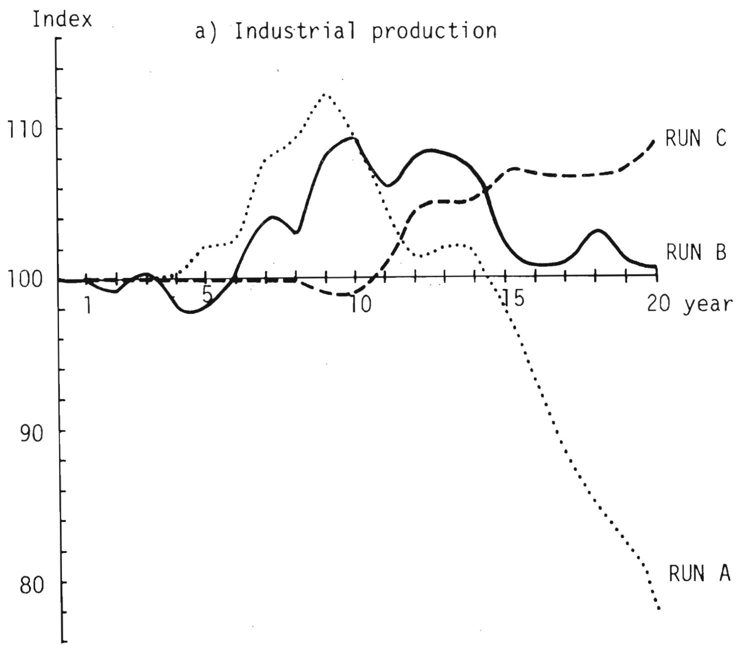
Diagram 1 traces the macro activity effects on industrial output and employment.

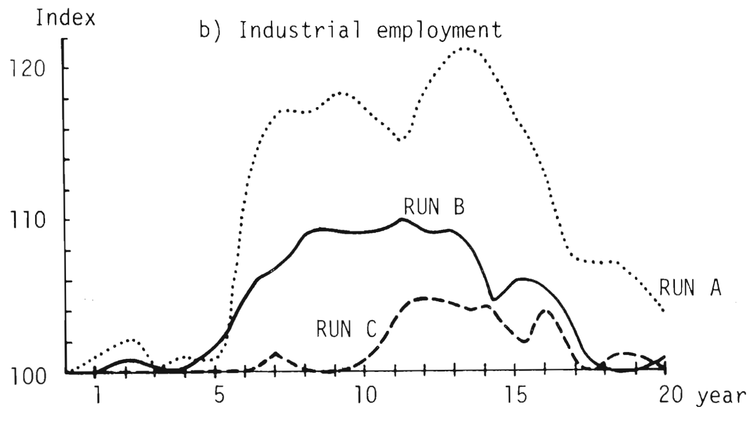
The first, 40 percent case is an induced change of the drastic kind. It spins off a positive (production and employment) effect of the expected kind in the beginning. However, after 10 years the multiplier-accelerator mechanisms at work from micro-to-macro and back again start to reverse themselves and production levels come down dramatically. For the first five years a small overall expansion effect in output and investment (not shown) is recorded. Over the 20 year period it is negative. Only the raw material sector has benefitted. The tendency towards relative decline is still there at the end of the simulation. Previous experience (from runs longer than 20 years) of the properties of the entire system tells us that production levels will not stabilize and start to grow again until employment has been trimmed down enough to restore profit margins and investment incentives. This will take more time since the employment effect is still positive after 20 years, and profit margins are on their way down, indicating a dramatic drop in productivity.

The 10 percent exogenous increase in raw material export prices gives a similar long term time profile, however, without the long term negative effects. The initial total production effect is negative (in sectors other than raw materials. See next section). The ensuing growth impulse, even though somewhat later, is equally strong and more enduring. It is still positive at the 20 year horizon and (NB!) the initial employment effect is just about nullified by then, suggesting a long run positive productivity effect.

However, when we substitute a 10 percent exogenous increase in potential output in the raw material sector for the exogenous price effect, the long term macro development changes. Essentially the two 10 percent changes mean the same to firms in the

Diagram 1





Note: The index measures the respective levels in the experiment in percent of the corresponding levels in a reference case.

For identification of RUN A, B and C, see Table 1.

Table 1 A-E. Effects on subindustry growth patterns from a very strong and a moderately strong (+40 and +10 percent) relative price change in the raw material sector and an exogenous productivity improvement (+10 percent) in the same sector

Identification: RUN A; foreign price up 40% 2nd year in RAW and maintained 20 years.

RUN B; ditto 10%

RUN C; Potential output up 10% 2nd year in RAW and difference maintained 20 years

Note 1: All comparisons are made vis-a-vis a reference case without the indicated, ceteris paribus, A, B and C changes, respectively.

Note 2: All tables except E give effects in percentage points per annum.

	20 years			First 5 years		
	A	B	C	A	B	C
A. Industrial production, percent per annum. Differences						
RAW (1)	4.8	5.8	-0.3(!)	6.4	0.8	2.9(!)
IMED (2)	-0.2	0.9	-0.4	0.8	-0.7	-0.5
DUR (3)	-3.1	-0.1	0.7	-0.5	-1.7	-1.3
CONS (4)	-0.5	0.2	0.3	0.3	-0.5	-0.3
TOT	-1.3	0	0.4	0.5	-1.2	-0.7
B. Labor productivity						
(1)	0.3	-0.7(!)	0.2(!)	1.1	-1.8	0
(2)	1.5	0.4	0	0.3	-0.4	0
(3)	-3.3	0.1	0.5	0.4	0	0
(4)	1.4	0	-0.1	0.4	0.1	0
TOT	-1.3	0.1	0.3	0.5	-0.4	0

Cont.

Table 1, cont.

	20 years			First 5 years		
	A	B	C	A	B	C
C. Wholesale (domestic) prices						
(1)	1.7	1.5	1.0	4.1	3.5	1.5
(2)	-0.5	-0.2	-0.1	-0.3	0.7	0.8
(3)	0.1	0	0.1	-0.1	-0.3	-0.2
(4)	0.1	0.3	0.3	0	-0.2	-0.3
TOT	0.4	0.2	0.2	0.9	0.1	0
D. Wages in industry						
(1)	2.0	1.3	0.5	2.4	-0.2	0.1
(2)	1.1	0.7	0.8	1.0	-0.1	0.1
(3)	1.8	1.5	1.4	0.7	-0.7	0.1
(4)	1.9	3.4	3.5	0.6	0	0.2
TOT	1.7	1.7	1.6	1.0	-0.5	0
E. Profit margins, index 100 = reference case						
(1)	118	96(!)	100	153	112(!)	100
(2)	87	96	100	99	100	100
(3)	45	101	98	89	102	100
(4)	94	100	99	99	101	100
TOT	76	99	99	107	103	100

Sector 1 (RAW) = Raw material production sector
2 (IMED) = Intermediate goods production sector
3 (DUR) = Durable consumption and investment goods sector
4 (CON) = Other goods production sector (for final consumption)

raw material sector in terms of potential profits. A conventional profit maximizing firm with full knowledge of what happened would have responded identically to the two changes. Not so here. The price change operates through external information gathering and on the interpretation sensors of the firm through expectations. Especially the 40 percent price change, but also the 10 percent change, throws previous interpretive mechanisms out of balance for a while and creates expectational mistakes. The increase in potential output is an internal, albeit exogenous, change. It creates a productivity reserve that is not made use of until needed to meet profit targets. That need does not arise for a while (in the simulation). Neither does this change disturb the market information system of the firms. I would argue that this "asymmetric" response pattern of firms is a highly realistic feature of business life.[1]

Hence under the technology shift short term growth performance takes time to improve but speeds up and is still on its way up at the 20 year horizon. The employment effect is only temporarily positive, suggesting again that firms eventually make use of the productivity potential given them from above.

On the macro surface of it it seems as if a too strong relative price change (+40 price case) produces such long term disturbances to the economy as to be undesirable, even though the short term impacts in the affected sectors are positive.

Two post war experiences of the Swedish economy should be recalled here. First, the overall exogenous price shock on Swedish industry in 1973 was between 30 and 40 percent. Two devaluations and an enormous infusion of subsidies were needed in 1977 to prevent a drastic sequence of closedowns in large parts of the manufacturing sector, and as this is being finally edited we do not know to what extent these measures will

---

[1] See Eliasson [1976a].

result in a new round of second generation inflation problems. The above price hike experiments on the model have been designed without these countermeasures to dampen structural change, but our contention is that the model simulation describes quite well in principle what has happened. The other experience was the Korean boom in 1951 with a more than 50 percent average price increase, most of it affecting forest industries. Since the price hike was more isolated and (unlike in 1973 to 1977) was strongly reversed in 1952 and 1953 disturbances did not get an opportunity to accumulate in momentum and the negative, secondary effects were much smaller.

The "softer" price stimulus (+10 percent) definitely is to be preferred to the stronger alternative, but also this alternative seems to come second to a stimulus that does not bring disturbances into the market information, interpretation system of firms, but rather lets new potentials dawn upon decision makers when the "need" for them arises.

5. RESULTS - ALLOCATION EFFECTS

As expected the allocation effects are extremely strong in the case with a 40 percent step price increase in the raw material sector. The raw material sector sets off on a happy boom and we see no end to it on the 20 years horizon. The sector that suffers, and especially so when the downward twist sets in, is investment goods industries.

These structural changes are indirectly and endogenously induced. And the prime factor at work is the labor market wage arbitrage function. In order not to loose too many people to the strongly expanding and profitable firms in the raw material sector, other firms, not as lucky, have to increase their wages. Some firms cannot follow suit, especially when (investment) demand starts to taper off. They contract operations and/or reduce investment spending. A very strong flow of labor resources (net) from all other sectors to the raw material sector occurs. While the raw material sector employed 14 percent of industrial (all four sectors) employment at the beginning of the simulation it employed 27 percent at the end, after 20 years. Indeed so strong and so fast has been the reallocation of labor that the ensuing wage drift has brought disturbances into the labor market, causing misinterpretation of price signals that has driven down profit margins in the three non-raw material sectors much below what would have been the case with a slower change.

It is of interest to note from table D how efficiently the labor market transmits the original price-wage effect in the raw material sector to other sectors. There is some spread in wage changes between sectors for the first five years, with relatively higher increases in the durable goods and consumption goods sectors, induced to grow by investment demand from all firms and consumer demand from households (the expansion phase of the multiplier-accelerator). Over the 20 year period, however, wage change is practically equal in all sectors.

Not so price change and productivity change (NB negative in the long run!) producing a tremendous dispersion in profit performance between sectors and firms. The direction of the effects are as expected. There is, however and unfortunately, no evidence around to assess the relevance of the magnitudes of the effects simulated.

The general character of the results are preserved for the softer 10 percent exogenous price change. As before, the change is large enough to distort the market price signalling system. The magnitude of the effects are much smaller. There are, however, some significant differences.

First and foremost, there is no long run "catastrophic" effect in the investment goods industries when the multiplier accelerator mechanisms go into reverse. By and large, however, the raw material sector increases its size measured in output substantially relative to the other three sectors.

The same pattern holds for labor productivity with the difference that the raw material sector takes out part of its exogenous price windfall in the form of a slackening of productivity performance.[1]

Another interesting structural response is that the derived demand for labor from the expanding raw material sector is no longer strong enough to even out wage change as efficiently as in the +40 case. This is, of course, part of the reason for an equally soft profit margin effect. In fact, even though the short term effect is strongly positive, the long term profit effect is negative in the raw material sector -- due to over-

---

[1] Also a highly realistic response (see Eliasson [1976a]). Also cf Carlsson [1972] who reports that productivity performance (in a technical sense) had been increasing fastest in sectors having a hard time while e.g. pulp & paper industries that at the time (1967) at least were thriving on an abundant raw material base by no means displayed a superior productivity ranking.

optimism and overexpansion.[1]

The 10 percent increase in the productivity potential in the raw material sector finally has a much softer structural as well as macro impact.

The initial (first 5 year) expansion draws resources away from other sectors. In the long (20 year) run, however, there is no real relative change in sizes between the four sectors. Neither is relative profit performance more than marginally affected.

One interesting feature is worth noticing, however. In the two first, price induced simulations, the foreign price change was "duly" transmitted through the economy and ended up in full in the consumer price index.

In the case with an exogenous increase in productivity the initial expansion in the raw material sector means that more people are needed and to get them raw material firms can pay roughly as much more as in the case with a 10 percent price hike without lowering their profit margins. This wage drift is transmitted through the entire economy to other sectors (costpush, perhaps) and to households (demand pull). The final outcome is a long run increase in the wholesale price level, although not as large as in the other two runs. If long run, stable growth is desired, of a kind that does not build up disequilibria that force a reversal after some time and (NB!) that is not associated with excessive inflation -- then the potential output hike is to be preferred to the price hike. Isn't this what Sweden benefitted from during the late 50ies and most of the 60ies between the Korean boom and the oil crisis? A price hike case somewhere in between the two price experiments would probably quite well illustrate the situation we are currently suffering from and what Sweden went through in the early 50ies.

---

[1] The 40 percent price hike in case A is so strong as to preserve a positive long term profit margin effect despite substantial overinvestment.

One concluding word about how exactly these results relate to the concept of international competitive advantage discussed earlier is now in place. Our experiments treat both the price hikes and the productivity shift as an initial improvement in the competitive position of our model economy. The price hike corresponds to an improved market position vis-a-vis the rest of the world, the productivity increase to the discovery of a new, non-imitable production technique or raw material resource -- both without effort (investment) on the part of the model economy. The experiments show that if doused too suddenly by too generous benefits from above, firm decision makers get confused, make inefficient decisions and the whole economy may eventually suffer. One may argue that a global price hike in one particular product should have a detrimental impact on that sector if it does not exhibit a comparative advantage in that particular kind of production. More efficient producers would respond by expanding production even more and check the price increase or even drive it back. The less competitive producers would find themselves with an inflated cost structure because of the initial price hike and an even worse competitive position. The experimental design rules out this possibility[1] by assumption. The point was to demonstrate that an initial improvement in the competitive position of a sector may carry reverse long term implications -- a possibility assumed away in most economic model building -- but not in reality.

---

[1] It would have to be engineered exogenous through the foreign price assumption.

BIBLIOGRAPHY

Albrecht, J, 1978, Production Frontiers of Individual Firms in Swedish Manufacturing 1975 and 1976, This conference volume.

Bentzel, R, 1978, A Vintage Model of Swedish Economic Growth from 1870 to 1975, This conference volume.

Carlsson, B, 1972, The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries 1968, The Swedish Journal of Economics 1972:4.

Ferguson, C E, 1975, The Neoclassical Theory of Production & Distribution, Cambridge University Press, Cambridge.

Pratten, C, 1976, A Comparison of the Performance of Swedish and U.K. Companies, Cambridge University Press, Cambridge.

Eliasson, G, 1972, International Competitiveness - An Empirical Analysis of Swedish Manufacturing, Federation of Swedish Industries, Economic Research Reports, B3. Stockholm.

- 1976a, Business Economic Planning - Theory, Practice and Comparison, Wiley, London.

- 1976b, A Micro-Macro Interactive Simulation Model of the Swedish Economy. IUI Working paper No 7. IUI, Stockholm.

- 1977, Competition and Market Processes in a Simulation Model of the Swedish Economy, American Economic Review, Vol. 67, No 1.

1978a, A Micro Simulation Model of National Economy in Eliasson (ed), A Micro-to-Macro Model of the Swedish Economy. IUI Conference Reports 1978:1. Stockholm

- 1978b, Experiments with Fiscal Policy Parameters on a Micro to Macro Model of the Swedish Economy; Paper prepared for the Conference on Micro Economic Simulation Models for the Analysis of Public Policy, The National Academy of Sciencies, Washington DC, March 1978.

CHOICE OF TECHNOLOGY IN THE CEMENT INDUSTRY – A COMPARISON OF THE UNITED STATES AND SWEDEN*

Bo Carlsson
IUI, Stockholm

1. INTRODUCTION

In recent years, a number of studies have shown that there are large international differences in energy consumption per unit of output in many industries. There are several reasons why such differences arise: the output mix varies, even within industries; the choice of technology varies; and input combinations differ even if the same technology is used.

It is natural for an economist to suppose that a large share of these differences can be explained by long-run international differences in relative factor prices. This was shown to be true, for example, in a recent study which compared the composition of industrial output and the use of energy in industry in the United States, Sweden and West Germany.[1] But it is obvious that there are many factors besides relative prices which also play an important role.

It is the purpose of the present paper to provide a more complete framework and, within this framework, to explain why the choice of technology varies internationally. Obviously, this kind of study requires rather detailed analysis and it is necessary,

---

[1] B. Carlsson, "Relativprisutvecklingen på energi och dess betydelse för energiåtgång, branschstruktur och teknologival i en internationell jämförelse" (The Development of Relative Energy Prices and Its Impact on Energy Use, Industrial Structure and Choice of Technology; An International Comparison). Appendix 12 to the report to the Energy Commission by the Expert Group on Policy Instruments. DS I 1977:17, Stockholm 1977. Also published by IUI as Booklet no. 83.

* I would like to thank Cementa AB, The Portland Cement Association and several U.S. cement firms for their generous assistance with information and advice. I would also like to thank the participants in the seminar on "Production, Technology and Industrial Structure" as well as professor Richard R. Nelson for helpful comments on earlier versions of this paper.

therefore, to focus on a particular sector and even on a single process.

For several reasons the cement manufacturing process has been chosen for this study: The output is homogeneous; the production process is relatively uncomplicated and separable from other processes; and it is known from the start that the choice of production techniques has been very different in various countries, at least up until recently. In addition, cement manufacturing is one of the most energy consuming processes in the whole of manufacturing industry.

As indicated in table 1, there are at least five types of processes used in cement production. The differences among them will be explained below. The purpose of the table is merely to show that even in an extremely homogeneous and capital intensive industry, the choice of technology may vary substantially among countries. The question with which we are concerned is why different choices are made. For reasons having to do with data availability, the analysis will be limited to a comparison of the United States and Sweden.

Table 1. International Differences in the Distribution of Cement Manufacturing Capacity by Process

Country		Wet process %	Semi-dry process %	Dry process, %			Shaft %
				Total	Long dry	Suspension preheater	
United States	(1976)	55	-	45	29	16	0
Sweden	(1975)	56	8	36	-	36	0
West Germany	(1974)	5	26	66	..	..	3
United Kingdom	(1974)	69	16	15	..	..	0
Italy	(1974)	13	46	40	..	..	1

Sources: Portland Cement Association, U.S. Portland Cement Industry: Plant Information Summary, December 31, 1976;
Cementa AB;
Gordian Associates, Industrial International Data Base, The Cement Industry, NATO/CCMS-46. New York.
Energy Research and Development Administration, 1976, p. 37.

Section 2 describes the cement manufacturing process and provides a brief history of the technological development of the industry. Section 3 brings out some salient features of the industry and how they differ between Sweden and the United States. This analysis is based largely on interviews conducted by the author during the Spring of 1977 in both the Unites States and Sweden. In Section 4, some investment cost calculations for both wet and dry kilns using price data for 1970 and 1975 will be presented. Section 5 discusses the differences between actual and theoretical costs of wet and dry process plants and section 6 analyzes the reasons for the delayed introduction of the suspension preheater process. Section 7 concludes the study.

2. THE CEMENT MANUFACTURING PROCESS - DESCRIPTION AND BRIEF HISTORY

The raw material for cement production consists mainly of limestone which is crushed and then ground into a fine powder. In the dry cement manufacturing process, the powder is fed directly into a kiln where it is calcined (burned) to form clinker. In the wet process, water is added to form a slurry which is then fed into the kiln. The kiln is essentially a huge cylindrical steel rotary tube lined with firebrick. Some kilns have a diameter up to 8 meters and are up to 230 meters long -- longer than the height of a 70-story building. The kiln axis is slightly inclined, and the raw material (either slurry or dry) is fed into the upper end. At the lower end is an intensely hot flame which provides a temperature zone of about 1500^{o} C by the precisely controlled burning of coal, oil or natural gas under forced draft conditions.[1]

[1] Energy Conservation Potential in the Cement Industry, Conservation Paper number 26, prepared by the Portland Cement Association for the Federal Energy Administration. June 1975 (Springfield,Va.: U.S. Department of Commerce, National Technical Information Service, PB-245 159), p.1.

After the clinker is cooled, it is ground with 4-6 % gypsum into cement.

The earliest cement kilns were dry process but of a different type (vertical shaft kilns) than the modern ones. In the early 1900's, long rotary horizontal kilns began to be introduced. Because of the relative ease of grinding and homogenizing the raw materials under wet conditions, the wet process came to dominate. The drawback of the method, however, is that it is much more fuel consuming than the dry process, since the water added in the raw mill must be dried away before calcination can take place.

In 1927, a semi-dry process was patented in Germany. It was named the Lepol process (acronym for the inventor, Lellep and the equipment manufacturer, Polysius).[1] The basic principle of the process is to use the exhaust gases from the kiln for drying and preheating the raw materials before inserting them into the kiln. Thus, the main advantage is energy saving. The process became popular in Europe but was hardly used at all in the United States.

In 1933, yet a new type of dry process cement kiln was patented in Czechoslovakia. Then World War II intervened, but after the war the patent was acquired by a German equipment manufacturer, and the first installation was made in 1950 in Germany.[2] In a conventional dry kiln, three sub-processes take place simultaneously. At the upper end, where the materials are fed into the kiln, preheating takes place. In the middle, calcining occurs, and at the lower end the final burning. Since only a fraction of the raw materials on the rotating kiln wall is exposed to the hot air,

---

[1] S. Mängel, Technischer Fortschritt, Wachstum und Konzentration in der Deutschen Zementindustrie. Doctoral dissertation. 1972, p. 24.

[2] Hoke,M. Garrett, "The Potential Promise - Prospects and Pitfalls in Energy Conservation by the U.S. Cement Industry", in Proceedings of the FEA-PCA Seminar on Energy Management in the Cement Industry. Federal Energy Administration Conservation Paper Number 47, FEA/D-76/091, p. 268.

the heat exchange is very inefficient and the fuel consumption therefore high. Also, since the sub-processes require different temperatures, it is difficult to optimize the temperature throughout the kiln, and different scale factors seem to apply. A number of interviews conducted by the author have indicated that difficult operational problems arise in long conventional kilns as the scale is expanded.

The essence of the new kiln is that it separates the preheating from the other sub-processes which take place in a conventional kiln. Preheating of the materials takes place in cyclones where hot air from the kiln is blown in the opposite direction to that of the powder, with the result that the powder is temporarily suspended in the air. This provides a much more efficient heat exchange between the air and the materials than can be achieved in a kiln and the amount of energy required is therefore reduced very significantly.

In recent years, Japanese firms with license rights on West German suspension preheaters have developed an auxiliary burner system in the preheater, so that not only preheating but also up to 95% of the calcination takes place before the feed enters the kiln. In such a precalcining system both the length and the diameter of the kiln can be further reduced, and energy consumption may also be slightly reduced. But the main advantage of the precalcing system may lie in its ability to deal with some operational problems encountered in suspension preheater systems.[1] The first precalcining system was developed in Japan in 1966.[2] The process has already been introduced in the United States (1976) and is currently being introduced in Sweden.

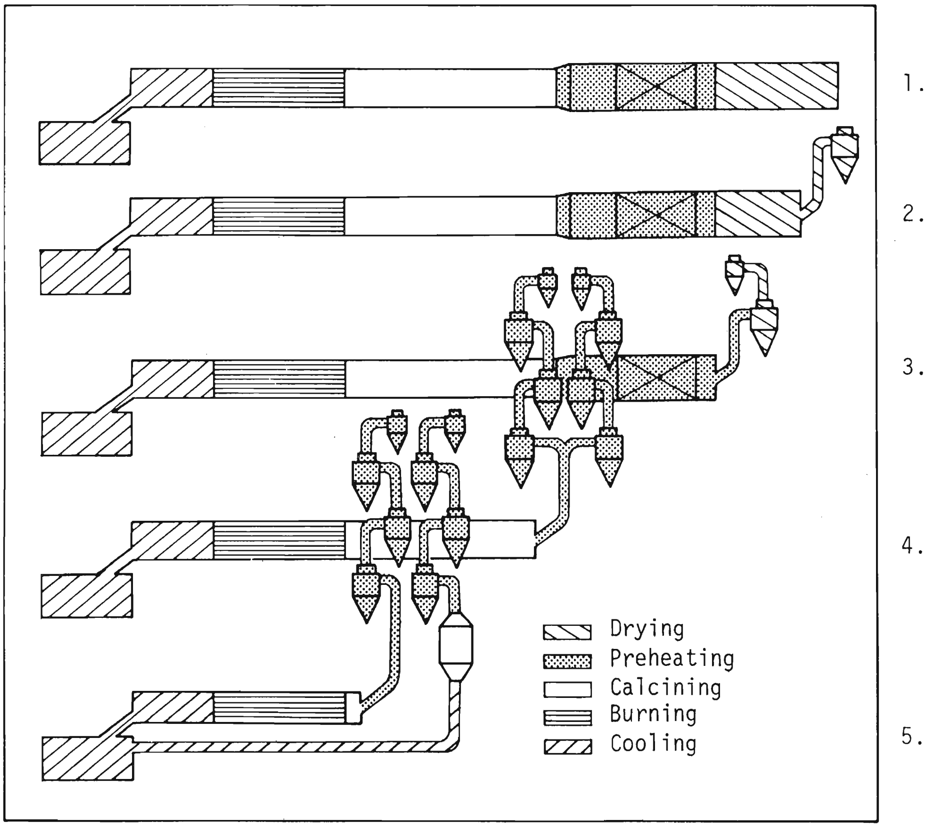
The development of cement production technology over the past 30 years is illustrated in figure 1. Until 1950, conventional long

---

[1] Gordian Associates, Industrial International Data Base, The Cement Industry. NATO/CCMS-46. New York: Energy Research and Development Administration. 1976, p.14.

[2] FEA-PCA Proceedings, p.267.

Figure 1. Technical Change in Cement Kilns



1. Conventional long (dry) kiln
2. Dry kiln with 1-stage preheater
3. Dry kiln with 2-stage preheater
4. Dry kiln with 4-stage preheater
5. Dry kiln with 4-stage preheater and precalciner

Source: FLS-newsfront. F.L. Smidth. Copenhagen 12.1974.

kilns were used. With the arrival of cyclone preheaters, the length and diameter of the kiln could be substantially reduced. In order to produce 1 225 tons in 1950, a kiln of 143 meters and 4.8 meters' diameter was required. In the 1970's, a kiln of 63 meters and a diameter of 4.2 meters could produce the same output.[1]

With the preheating of the materials taking place outside the kiln, the length and the diameter of the kiln can be substantially reduced for the same capacity. This, in turn, means a (theoretical) saving in capital cost, since preheater cyclones are cheaper to build and install than the additional kiln section which would otherwise be required. Alternatively, for the same capital cost, much larger capacity can be obtained. Since the number of people required to operate the kiln is about the same, no matter what size and type of the kiln, the suspension preheater process also offers substantial labor saving.

3. INDUSTRY CHARACTERISTICS

3.1 Post-War Development in the U.S. and Swedish Cement Industries

Output and Capacity

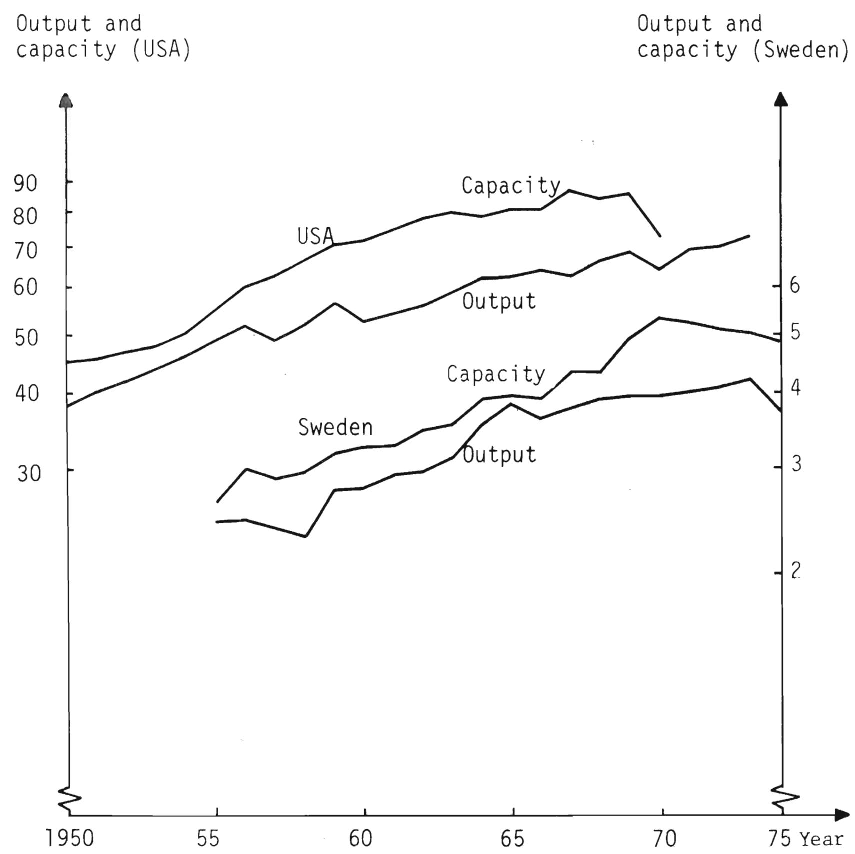
Cement production grew in the United States at a rate of 2.9% per year 1950-73 and at 2.3% per year in Sweden 1950-75. In both countries cement production grew less rapidly than total manufacturing output. However, as can be seen in figure 2, the growth rate has been fairly constant over the whole period in the United States while it was high in Sweden up to 1965 and has since stagnated. It is shown also in figure 2 that from the mid-1950's U.S. production capacity increased much faster than output. This re-

---

[1] H.R. Norbom, "Wet or Dry Process Kiln for your New Installation?" Rock Products, Vol.77, No.5 (May 1974), pp.92-93.

Figure 2. Actual Output and Production Capacity in the Cement Industry in the United States and Sweden 1950-1975

Million metric tons



Sources: Energy Conservation Potential ..., op.cit., p. [illegible] (Production obtained as U.S. consumption minus imports);

FEA-PCA Proceedings ..., op.cit., p. 43;

Cementa AB

sulted in considerable overcapacity which persisted until the end of the 1960's. In Sweden the capacity utilization has been higher on the average than in the United States (85% vs. 81%), but it has fallen drastically after 1968 when the increases in the demand for cement fell far short of the capacity increases.

Kiln and Plant Size

In spite of the fact that Swedish cement capacity in 1975 was only about 6% of U.S. capacity, table 2 shows that Swedish cement kilns are larger, on the average, than American ones. This is true not only at the present time; they have also been larger in each time period (with two exceptions: 1936-40 and 1961-65).

In table 3 the size and age structure of cement kilns and their distribution by process in the United States and Sweden are shown. The majority of cement kilns and more than half of cement capacity in both countries are still wet process. However, in Sweden no wet kilns have been installed since 1967, while in the United States the last wet kilns were put in in 1975.

Other major differences between Sweden and the United States arise through the differences in the size and distribution by process of dry kilns. Swedish dry kilns are 50% larger than U.S. dry kilns. This has to do with the fact that over 80% of the Swedish dry kiln capacity is in suspension preheaters, whereas in the United States the corresponding figure is only 35%. No long dry kilns at all exist in Sweden, but there are two semi-dry kilns which are scheduled to be scrapped in 1978. It is also noteworthy that the two Swedish SP kilns built in 1969-70 are larger than the five American SP kilns built in 1976.

It is also true that Swedish cement plants are larger than U.S. plants: the average Swedish plant had a capacity in 1975 of 725 000 tons of cement, while the average American plant had a capacity in 1976 of 545 000 tons.[1]

---

[1] "Tons" refers to metric tons throughout unless otherwise stated; 1 metric ton = 1.1023 short tons.

Table 2. Size and Age Structure of Kilns in the Cement Industry in the United States (1976) and Sweden (1975)

	United States			Sweden		
Kiln age	No. kilns	Clinker capacity 1 000 metric tons	Average capacity per kiln 1 000 metric tons	No. kilns	Clinker capacity 1 000 metric tons	Average capacity per kiln 1 000 metric tons
1976	6	2 800	467	0	-	-
1971-1975	34	13 766	405	0	-	-
1966-1970	34	11 606	341	3	1 732	577
1961-1965	47	14 272	304	4	858	215
1956-1960	82	16 336	199	1	214	214
1951-1955	59	8 930	151	4	584	195
1946-1950	36	4 757	132	3	584	195
1941-1945	9	1 316	146	1	190	190
1936-1940	7	1 366	195	4	620	155
1931-1935	6	615	103	0	-	-
Before 1931	65	5 687	87	0	-	-
	385	81 451	212	20	4 993	250
Year of construction of average kilns						
based on number of kilns			1952		1953	
based on clinker capacity			1959		1953	
Share of clinker capacity in dry process, %			45		44	

Sources: PCA, U.S. Portland Cement Industry: Plant Information Summary. December 31, 1976; Cementa AB.

Table 3. Size and Age Structure of Cement Kilns and Distribution by Process in the United States (1976) and Sweden (1975)

	United States				Sweden			
	No. kilns	Clinker capacity 1 000 metric tons	Average kiln capacity 1 000 metric tons	Share of total capacity %	No. kilns	Clinker capacity 1 000 metric tons	Average kiln capacity 1 000 metric tons	Share of total capacity %
Wet process								
Total	214	44 750	209	55	13	2 796	215	56
1976	0	-	-		0	-	-	
1971-75	14	5 236	374		0	-	-	
1966-70	24	8 129	339		1	445	445	
Dry process	171	36 700	215	45	7	2 197	314	44
Long dry[a]								
Total	119	23 300	196	29	2	412	206	8
1976	1	240	240		0	-	-	
1971-75	2	917	459		0	-	-	
1966-70	9	3 098	344		0	-	-	
Suspension preheater								
Total	52	13 400	258	16	5	1 785	357	36
1976	5	2 560	512		0	-	-	
1971-75	18	7 612	423		0	-	-	
1966-70	1	379	379		2	1 287	644	
Total all processes	385	81 450	212	100	20	4 993	250	100

[a] Refers to semi-dry kilns in Sweden.

Sources: See table 2.

Labor Productivity

Even though both kilns and plants tend to be larger in Sweden than in the United States, labor productivity in the United States has remained higher than in Sweden throughout the period. Se figure 3. However, the labor productivity gap has narrowed from 49% difference in 1950 to 25% in 1974. On the other hand, it is also shown in figure 3 that the total wage cost per hour has increased far more rapidly in Sweden than in the U.S., so that in 1975 the Swedish wage rate exceeded the American one. Thus considering the labor productivity difference, the Swedish wage cost per ton of cement surpassed the U.S. wage cost in 1971 and was as 51% higher than the American wage cost in 1974. (See also figures 6 and 7 below.)

Energy Consumption

At the same time as labor productivity has increased in both countries, fuel consumption has also been reduced, as illustrated in figure 4. The reduction has been about 25% in the United States and 20% in Sweden, but the remaining difference is still very large. For comparison, the fuel consumption in West Germany during the same period is also shown in figure 4 and is found to be still lower than that in Sweden.

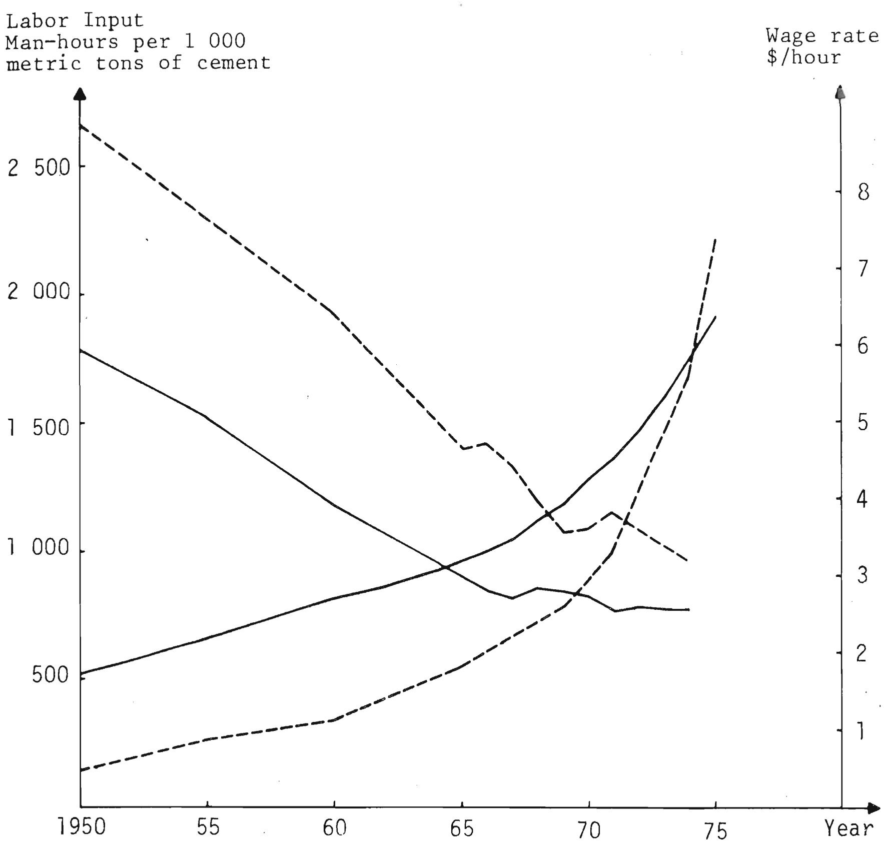
3.2 Overall Industry Characteristics

There are four characteristics of the cement industry which go a long way towards explaining the differences between the Swedish and the American cement industries observed above. These are economies of scale, high energy intensity, high transport costs, and homogeneous output.

a) Economies of Scale

The presence of economies of scale in the cement industry is illustrated in figure 5. There are substantial economies of scale in both wet and dry plants. The investment cost per ton of annual capacity is lower (at least theoretically) for dry than for wet

Figure 3. Development of wage rates and labor productivity in the cement industry 1950-75 in the United States and Sweden

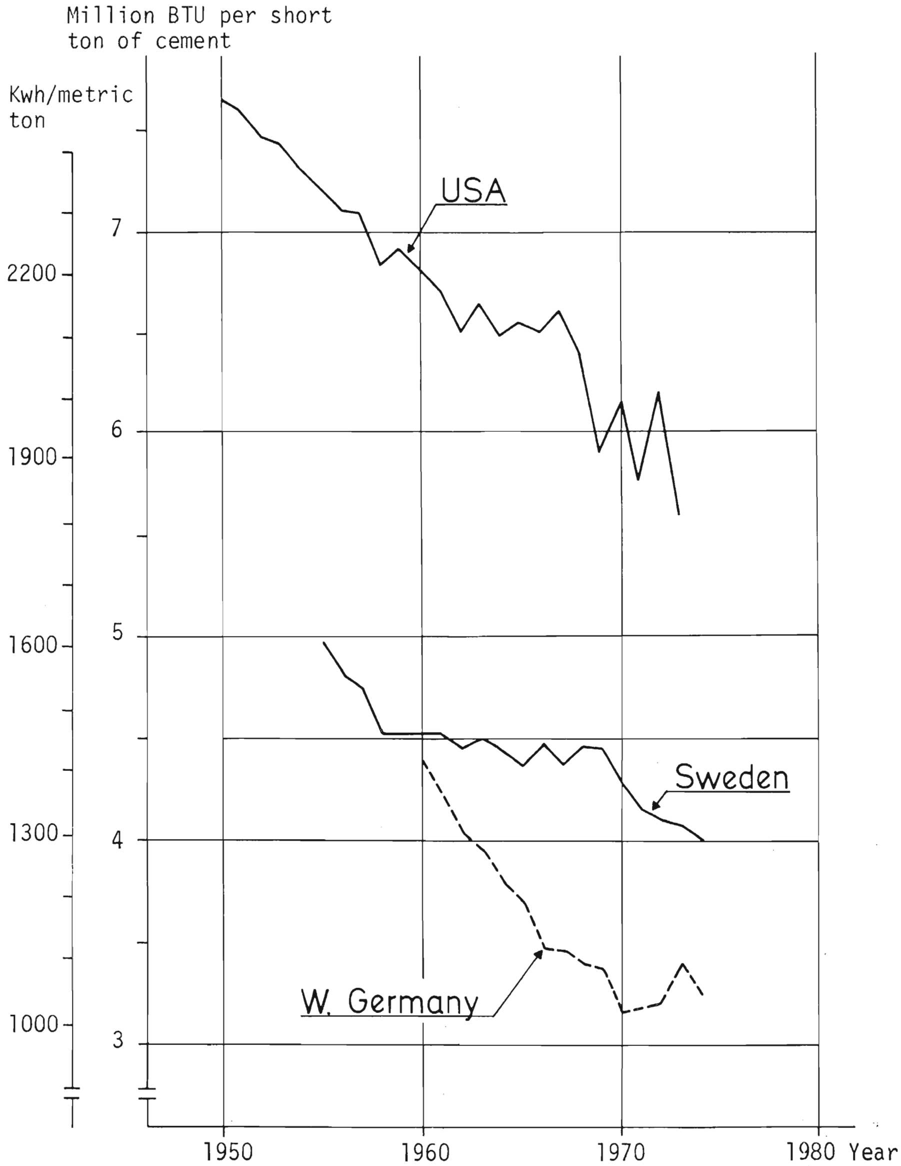


Note: U.S. figures include both direct and overhead labor. The Swedish figures have been made comparable in the following way: Administrative personnel are assumed to work the same number of hours as production workers, and the number of hours in these two categories have been added for the cement industry. The same assumption is made for employees in limestone quarries. Employment in limestone quarries has been obtained by assuming that the proportion of limestone quarry employees out of total quarry employment has remained at the 1973 level throughout. This was the only year for which separate data for limestone quarries were available.

Sources: Labor productivity:SOS, Industri for each year; FEA-PCA proceedings, op.cit, pp.25-27. Wage rate in manufacturing:Swedish Employers' Confederation, Direct and Total Wage Costs for Workers, Various issues. U.S. figures for 1950 and 1955 have been obtained by chaining together with data from Council of Economic Advisers, Economic Report of the President, January 1966 (Washington: USGPQ, 1966), p.243. Swedish hourly salaries 1950-73 have been obtained from SOS, Löner 1973, Part 2, and for 1974-75 from Allmän Månadsstatistik. Total wage costs have been obtained by adding fees and charges for social benefits according to information from the Swedish Employers' Confederation. The Wage rate is expressed in current prices. The Swedish figures have been converted into dollars using the average official exchange rate for each year.

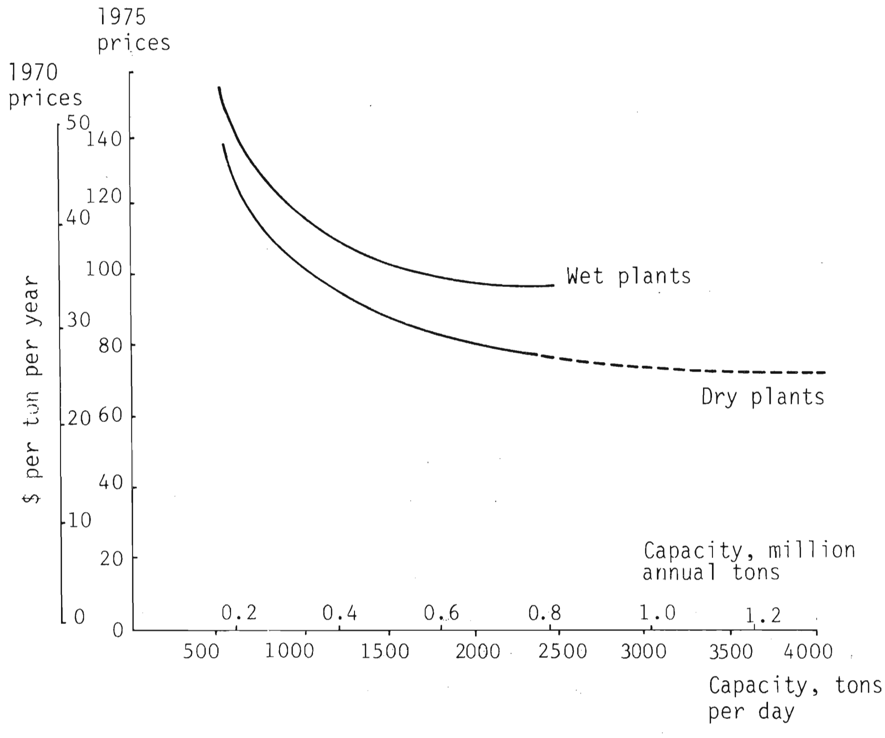
Figure 4. Fuel consumption in cement production in the United States, Sweden and West Germany 1950-74

Million BTU per short ton of cement



Sources: G.A. Schroth, "Grade Preheater Kiln Systems" in FEA-PCA Proceedings, op.cit., p. 253. Cementa AB.

Figure 5. Investment Costs in Wet and Dry Cement Plants, 1970 and 1975



Sources: K.T. Andersen, "Kiln Selection", in Proceedings of the FEA-PCA Seminar on Energy Management in the Cement Industry, Conservation Paper No. 47, 1975, p. 207.

S. Mängel, Technischer Fortschritt Wachstum und Konzentration in der Deutschen Zementindustrie, doctoral dissertation, pp. 47-48.

plants and continues to decrease beyond where the investment cost per ton in wet plants levels off.

b) High Energy Intensity

The cement industry is extremely energy intensive. In Sweden, the fuel and electricity cost has ranged between 29 and 41% of the value of sales during the period 1950-75. In the United States, the corresponding range was 19 to 28%. The energy cost has been higher than the labor cost throughout the period studied in both countries.[1] We will return later to the energy considerations in detail when discussing the choice of technology.

c) High Transport Costs

Because of the relatively low price per ton, the relative transport costs of cement are extremely high. This means that the cement industry is highly local in character, especially in regions without water transport facilities. It costs as much to transport a ton of cement 100 km by truck in Sweden as 600 km by small coastal shipping vessels or 2 000 km by large bulk carries.[2] Therefore, in order to utilize scale economies fully, cement plants must be located either in large metropolitan areas or on waterways.

Because of the high transport costs not only for the finished product but also, and even more so, for raw materials, the cement industry is forced to rely on local raw materials which may vary greatly in quality among locations. Thus, the moisture content and purity of the raw materials as well as their hardness and accessibility vary substantially among plants.

d) Homogeneous Output

Although the quality of cement can theoretically vary among plants and locations, most countries impose fairly stringent requirements

---

[1] See figures 6 and 7 below.

[2] B. Carlsson, "Industrins energiförbrukning 1974-80" (Industrial Energy Consumption 1974-80), Appendix 7 to <u>IUI:s långtidsbedömning 1976</u> (IUI's Medium Term Survey 1976), IUI. Stockholm, 1977, p.277.

which must be met by cement sold domestically. These requirements pertain to compressive strength, fineness, chemical composition, etc. They vary somewhat among countries, although the differences are not great among Western countries. It does seem, however, as though the U.S. specifications are more stringent in terms of both fineness and purity (esp. concerning the presence of alkalis) than those of West European countries.[1] The fact that U.S. cement is more finely ground essentially leads to slightly higher energy consumption than would otherwise be the case. The stricter alkali requirements may have more far-reaching consequences for the choice of technology, however, as will be shown below.

The presence of substantial economies of scale in combination with high transport costs has important implications for market structure. In Sweden, six out of a total of seven cement plants are located near water. This has made it possible to take advantage of scale economies in production. In addition, because of an extremely high degree of concentration (there is now only one domestic cement firm in Sweden after a merger in 1974), it has been possible to plan the production expansion in such a way that there is very little overlap geographically between plants. The primary reason why the Swedish government allowed the merger to go through was precisely the argument that this would facilitate achieving a more optimal industry plant structure, provided at the same time that there would be no tariff or other protection, and that the company would be subjected to price control. The product on capacity of the Swedish cement company is large enough to place it among the four largest U.S. firms.

By contrast, there were 52 cement companies in the United States in 1976, the largets of which had 6.7% of industry capacity. The four and eight largest firms accounted for 22.3 and 39.2% respectively.[2]

---

[1] Gordian Associates, <u>op.cit.</u>, p.39.

[2] Portland Cement Association, <u>op.cit.</u>, p.3.

The plants within the largest firms are also widely scattered geographically, making it difficult to concentrate production to one location without involving major changes in regional market shares. There were 162 plants in the U.S. in 1976. This large number can be explained by both geographical factors (population density, transport costs, large inland areas without access to water transport facilities, etc.) and historical factors (most plants were built when scale advantages were less pronounced in areas where cement was needed and local raw materials were available).

While the above factors explain the plant structure of the U.S. industry, the size structure of kilns may be regarded as the consequence of another but related set of factors. During the last fifteen years, kilns built in the United States have tended to be relatively small. Immediately after the Second World War there was a shortage of cement capacity in the United States which led to overinvestment in the 1950's and early 1960's. The resulting overcapacity seems to have lasted into the early 1970's, making it unattractive to invest in anything but replacements of old, inefficient facilities. Since replacing an old wet kiln by a suspension preheater system would involve replacing much of the raw material handling equipment as well, there is a certain minimum scale below which the capital cost would be prohibitive.

How can one explain the observed labor productivity differences, when there are no differences in the average age of kilns and the size factor should imply an advantage for Swedish producers? While it has not been possible within the framework of this investigation to penetrate this question, since it would require a very large set of data for each plant, at least <u>one</u> plant comparison has been made. See table 4, where an old wet process American plant is compared to a relatively new Swedish plant with one wet and one large dry kiln.

Table 4. Comparison of structure of employment in an American and a Swedish cement plant 1976

	American plant	Swedish plant
Production capacity, 1000 metric tons/year	270	820
Average age of kilns, years	51	10
Number and type of kilns	3 wet	1 wet,1 dry
Total number of employees	109	330
Hourly	73	254
Salaried	36	76
Potential labor productivity 1000 tons/employee/year[a]	2.5	2.5
Distribution of labor force, %		
Quarry	4	9
Raw grinding	6	4
Burning and cooling	7	5
Finished grinding	6	9
Laboratory	11	3
Packing and shipping	13	16
Mechanical maintenance	23	22
Electric maintenance	7	7
Yard + other	23	25

[a] At full capacity utilization.

Note: Administration and other overhead employment has been distributed on the various departments according to the number of production workers.

Sources: Firm interviews.

The Swedish plant is about three times as large as the American plant in terms of both capacity and employment, i.e. labor productivity is about the same (namely 2,500 tons/employee/year which works out to about 720 man-hours per 1 000 metric tons under the assumption of 1 800 working hours per year -- or 6 % higher than the U.S. national average in 1974 and 25 % higher than the Swedish average). The proportion of salaried employees is slightly higher in the American than in the Swedish plant. As far as the employment in various departments is concerned, the differences do not seem to be overwhelming. The fact that the Swedish plant has more than twice the employment of the American plant in the quarry has to do with the fact that the raw material is a soft marl which can simply be bulldozed in the American case and hard limestone which requires the use of explosives in the Swedish case. The lower Swedish shares in the raw grinding and burning and cooling departments as well as the laboratory may be due to the newer, larger and more fully automated equipment. This is true especially in the laboratory. Both plants have relatively high employment in the finished grinding and packing and shipping departments, since they are both versatile plants which produce a variety of types of cement in both bulk and bagged form. (Most plants in both Sweden and the U.S. produce only one type of cement which is sold only in bulk.) Differences in product mix may explain the differences which do exist in these departments. The remaining service departments have virtually the same shares of employment in both plants.

The conclusion which emerges from this comparison is that there seems to be no major difference in the structure of employment in these two plants other than in the quarry and in the laboratory. But perhaps no such difference should be expected, since labor productivity is the same in both plants. It is remarkable, however, that labor productivity is as high in an old American plant as in a relatively new, clearly above average, Swedish plant. It would be interesting to compare the American plant to a Swedish plant of the same size and age, but unfortunately that has not yet been possible. Visits by the author to a number of both Swedish and U.S. plants indicate that Swedish plants tend to have considerably larger employment in the quarry (due to raw material differences) and yard departments (due to differences in preferences regarding working conditions and the amount of services in terms of cafeterias, health care, etc., offered by the company). In the operative departments, there seem to be no major differences. But this can only be conjecture until a more thorough investigation is made.

3.3 Production Costs and Cement Prices

In view of the fact that there have been 50-60 cement companies in the United States throughout the period and only one or two in Sweden, one might expect the pressure of competition to have kept the price lower in the United States than in Sweden. A look at figures 6 and 7, however, will show that just the opposite has been true. The price of cement has been 13 to 63 % higher in the U.S. than in Sweden, the price difference being especially great around 1960.

Cost differences seem to explain only part of this difference. As shown in figures 6 and 7, the total variable cost (labor plus fuel and electricity) was higher in the U.S. until 1965 but has since been lower. The U.S. labor cost per ton of cement was substantially higher than the corresponding Swedish figures during the 1950's, approximately the same during the 1960's and early 1970's and then 20 % lower in the last few years due to extremely rapid Swedish wage increases, coupled with devaluation of the dollar. Swedish fuel costs per ton of cement were considerably higher than those in the United States in the 1950's, only slightly higher in the 1960's, rising again in the 1970's in relation to the U.S. fuel costs. Thus, even though the U.S. fuel consumption was about 40 % higher than the Swedish one throughout the period, the fuel costs were lower than in Sweden, primarily due to the availability of cheap domestic natural gas and coal. Sweden

Figure 6. Production Costs and Cement Prices in the United States 1950-74

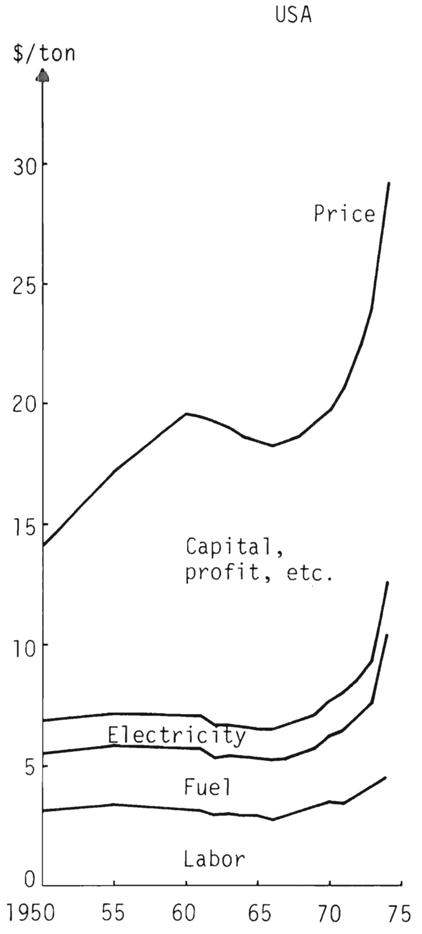
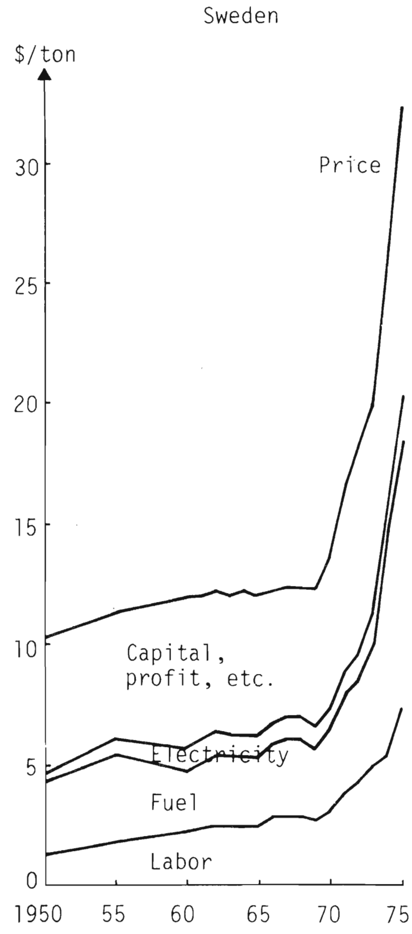


Figure 7. Production Costs and Cement Prices in Sweden 1950-74



Sources: See next page.

Figure 6 (Sources)

Cement price:	1950-70: FEA-PCA Proceedings ..., op.cit., p.43. 1971-74: U.S. Bureau of Mines, Minerals Yearbook 1974, Vol.1, p.283.
Electricity cost:	Electricity consumption: G.A. Schroth, op.cit., p. 236. Electricity price: Edison Electric Institute, Historical Statistics of the Electric Utility Industry, EEI Publication 62-69, New York, 1962, table 45. EEI, Statistical Yearbook of the Electric Utility Industry for 1975, EEI Publication No.76-51, New New York, 1976, table 60 S.
Fuel cost:	Total energy use: PCA, Conservation Potential .. op.cit., p.15. Distribution of energy consumption on fuel type: FEA-PCA Proceedings ..., op.cit., p.35. Price of coal: Minerals Yearbook, various issues. Price of gas: American Gas Association, Gas Facts, 1950, 1951, 1975, 1976, Arlington, Va. Price of oil: Platt's Oil Price Handbook and Oil Manac 1976, New York, McGraw-Hill Inc, 1976
Labor cost:	FEA-PCA Proceedings ..., op.cit., pp. 25-27.

Figure 7 (Sources)

Cement price:	SOS Industri, National Central Bureau of Statistics Stockholm, various issues.
Electricity cost:	Electricity consumption: Ibid. Electricity price: State Power Board.
Fuel cost:	Fuel consumption: SOS Industri Fuel prices: Svenska Petroleum Institutet, En bok om olja, Stockholm, SPI, 1971; Svenska Esso AB, Oljeåret 1975; SOS Utrikeshandel, various issues.
Labor cost:	Figure 3 in the present paper.

lacking both of these resources, had to import fuel and came to rely primarily on oil.

However, the availability in Sweden of cheap hydro power led to low electricity prices which show up in our calculation. Thus, the cost of electricity per ton of cement was only 1/3 of the U.S. electricity cost in 1950. In absolute terms, the cost difference was about the same throughout the period. Taken together, fuel and electricity costs have been roughly the same in both countries until 1971 when fuel costs began to rise in Sweden.

The overall conclusion one can draw from this price and cost comparison is that gross profit per ton of cement has been very substantially higher in the United States than in Sweden during the entire period. It has grown from \$ 7.18 per ton in 1950 to \$ 16.52 in 1974, while the corresponding Swedish figures are \$ 5.50 and \$ 9.77. Even if capital costs in the U.S. had been higher than in Sweden, which may have been the case but is relatively unlikely, it seems fair to conclude that net profits must have been considerable higher per ton in the U.S. than in Sweden over the whole 24-year period. It is apparent, however, that the overcapacity which existed in the U.S. cement market in the 1960's put a substantial downward pressure on prices and thereby profits. Given the general rate of inflation in the economy, the profits squeeze may have been serious in many companies by the early 1970's -- but worst in Sweden where the general rate of inflation has been higher than in the United States.

In order to put these results in some perspective, it can be mentioned that the Portland Cement Association has estimated that the investment cost of a new cement plant in the U.S. was \$ 88 per metric ton in 1974.[1] Assuming 20 years' depreciation and 15 % discount rate, the capital cost amounts to \$ 14 per ton in 1974 prices. This is only slightly less than the average 1974 gross profit per ton calculated above for the U.S. and over 40 %

---

[1] Energy Conservation Potential ....., op.cit., p. 19.

higher than the calculated Swedish gross profit. Although the development of investment costs per ton of cement over the last 25 years is not known, it is not likely that investment in the cement industry has been very profitable since 1960.[1]

4. WET vs. DRY PLANTS -- A THEORETICAL COST COMPARISON

It was argued earlier that all the major cost components are theoretically lower for preheater dry than for wet process kilns: the investment cost per ton of capacity is lower, and the labor and fuel costs per ton of output are also lower. But if this is true, also in practice, how is it possible that U.S. firms kept investing in wet kilns until 1975 and that the wet process share of total U.S. cement production increased until at least 1970? [2] How big are the cost differences between preheater dry and wet process kilns?

In order to gain some idea of the answer to this question, let us make a standardized cost calculation for a typical wet process and dry process installation in 1970 and then a similar comparison for 1975 (after the energy price changes), using aggregate national price data for both years. We will then report on the range of variation in actual costs and input requirements among individual plants obtained from interviews with cement firms in both Sweden and the United States.

In table 5 a comparison is made of the total cost of production in a new wet plant in the U.S. and Sweden to that of a new preheater dry plant, using average prices for both countries in 1970 and representative input requirements. The price assumptions are based on available national price averages for energy

---

[1] It is an interesting question for further research what the reasons are for the low profitability in Sweden and whether this is a general phenomenon.

[2] Energy Conservation Potential ..., op.cit., p. 12.

Table 5. Hypothetical Cost Comparison between Dry and Wet Process Cement Plants in the U.S. and Sweden, 1970

Cost item	Price per unit, $		Wet method, 600 000 tons/year: Requirement per ton of cement		Wet method, 600 000 tons/year: Cost, $/ton of cement		Dry method, 600 000 tons/year: Requirement per ton of cement		Dry method, 600 000 tons/year: Cost, $/ton of cement	
	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden
Coal	0.40	0.68	2.1 MBTU	0.0	0.84	0.0	1.40 MBTU	-	0.56	-
Natural gas	0.38	-	2.6 MBTU	0.0	0.99	0.0	1.75 MBTU	-	0.67	-
Fuel oil	0.49	0.60	0.5 MBTU	4.8	0.25	2.88	0.35 MBTU	3.1	0.17	1.86
Total fuel	0.40	0.60	5.2 MBTU	4.8	2.08	2.88	3.50 MBTU	3.1	1.40	1.86
Electric power	9.50	7.30	0.13 MWh	0.10	1.24	0.73	0.14 MWh	0.10	1.33	0.73
Total energy					3.32	3.61			2.73	2.59
Other variable costs	1.00	1.00	1.50 $	1.50	1.50	1.50	1.00 $	1.00	1.50	1.50
Total variable costs					4.82	5.11			4.23	4.09
Labor	4.25	3.00	0.45 hours	0.54	1.91	1.62	0.45 hours	0.54	1.91	1.62
Capital	1.00	1.00	5.51 $	5.51	5.51	5.51	4.71 $	4.71	4.71	4.71
Total production cost					12.24	12.24			10.85	10.42
Cement price					19.43	13.68			19.46	13.68

Note: MBTU = Million British Thermal Units. 1 MBTU = 293 kWh.
Sources: See text.

and labor in the stone, clay and glass products industry. The investment cost per annual ton of plant capacity has been obtained from a German study. See figure 5.

The investment cost assumptions made for 1970 in table 5 are \$ 34.50 per ton of annual capacity for the wet plant and \$ 29.50 for the dry plant. With a 15 % discount rate and 20 years' depreciation this amounts to a capital cost per ton produced of \$ 5.51 and \$ 4.71, respectively.

As far as labor requirements are concerned, it is assumed that both plants would require 150 employees in the U.S. and 180 in Sweden, with an average of 1 800 hours worked per year.

The energy consumption (both fuel and electricity) is assumed to be that of best practice plants of the respective kind in both countries. As indicated in the table, the American energy consumption figures are somewhat higher than the Swedish ones in view of the existing differences in operating practices and product specifications. The distribution on types of fuel corresponds to the averages for the cement industry in each country in 1970.

In spite of the large differences in both prices and input requirements, the overall cost picture turns out to be remarkably similar in the two countries both for the total costs and for the major cost components. The wet method turns out to be about 15 % (or about \$ 1.50) more expensive per ton produced than the dry process in both countries. But in Sweden the existing price of cement permitted a profit of only \$ 1.50 per ton with the wet method, while the profit margin was \$ 3 per ton with the dry method. Due to the considerably higher prices in the U.S., <u>both</u> methods were highly profitable, although the profit margin was about \$ 1.50 per ton larger for the dry process.

In table 6, the same comparison is made using 1975 prices and input requirements. Relative prices have changed considerably, and the distribution on fuel types has changed in line with present trends. Thus, both fuel prices and investment costs have approximately trebled, while the wage rate increased by "only" 140 % in Sweden and by 50 % in the U.S. In this manner the costs of cement production more than doubled in both countries. The dry method is still considerably cheaper than the wet process, but the absolute cost difference has trebled. At the same time the cement price development has been such that it is no longer possible to cover the costs of production in newly built wet plants even in the United States. On the other hand, the dry method does not seem very profitable either. But this is probably due largely to the excess supply of cement in the world market during the last several years.

5. ACTUAL vs THEORETICAL COST DIFFERENCES BETWEEN WET AND DRY PLANTS

Thus, if we look at national averages, it is easy to see why no wet kilns have been built in Sweden since 1967 nor in the United States since 1975. But if our cost calculations are at least roughly representative of the industry, there still remains a good bit to be explained. If firms are rational, and if a higher profit is regarded as more desirable than a lower profit, then how can we explain why wet plants continued to be built for so long in both countries? Perhaps the national averages gloss over differences among plants which would explain this seemingly erratic or irrational behavior? Perhaps the cost differences between wet and dry plants are not as great in practice as in theory?

In May-June, 1977, the author of this study carried out a number of interviews with representatives of cement firms in both Sweden

Table 6. Hypothetical Cost Comparison between Dry and Wet Process Cement Plants in the U.S. and Sweden, 1975

Cost item	Price per unit, $		Wet method, 600 000 tons/year: Requirement per ton of cement		Wet method, 600 000 tons/year: Cost, $/ton of cement		Dry method, 600 000 tons/year: Requirement per ton of cement		Dry method, 600 000 tons/year: Cost, $/ton of cement	
	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden	U.S.	Sweden
Coal	1.12	1.71	4.05 MBTU	2.40	4.54	4.10	2.73 MBTU	1.55	3.06	2.65
Natural gas	0.99	-	0.73 MBTU	-	0.72	-	0.49 MBTU	-	0.49	-
Fuel oil	1.93	2.09	0.42 MBTU	2.40	0.81	5.02	0.28 MBTU	1.55	0.54	3.24
Total fuel	1.17	-	5.20 MBTU	4.80	6.07	9.12	3.50 MBTU	3.10	4.09	5.89
Electric power	19.20	11.80	0.13 MBTU	0.10	2.50	1.18	0.14 MBTU	0.10	2.69	1.18
Total energy					8.57	10.30			6.78	7.07
Other variable costs	1.00	1.00	1.50 $	1.50	1.50	1.50	1.50 $	1.50	1.50	1.50
Total variable costs					10.07	11.80			8.28	8.57
Labor	6.50	7.20	0.45 hours	0.54	2.93	3.89	0.45 hours	0.54	2.93	3.89
Capital	1.00	1.00	15.60 $	15.60	15.60	15.60	14.11 $	14.11	14.11	14.11
Total production cost					28.60	31.29			23.32	26.57
Cement price					26.52	25.40			26.52	25.40

Note: MBTU = Million British Thermal Units. 1 MBTU = 293 kWh.

Sources: See text.

and the United States, major equipment manufacturers, a consultant firm, and the industry's branch organization in the United States, the Portland Cement Association. Data were gathered for a large number of plants in both countries. Emphasis was put on plants built in the late 1960's and mid-1970's -- investment costs, operating and price data, and especially the judgements made in connection with major investments and the reasons for building the particular type and size of plant.

Looking first at the empirical evidence concerning energy, a glance at table 7 will show quite clearly that suspension preheater and precalciner systems do offer considerable energy savings in comparison with both wet and conventional (long) dry systems. Converted into cost terms by using 1976 U.S. energy prices, the difference in energy consumption between preheater dry and wet process plants amounts to \$ 2.00-2.50 per ton of cement. The savings are greatest in fuels, whereas at least in U.S. operations the electricity consumption is higher in preheater dry than in wet systems. In both dry and wet systems, the Swedish plants seem to be more energy efficient.

The prices quoted for coal in 1977 ranged from \$ 0.84/MBTU (\$ 22 per metric ton) to \$ 1.75/MBTU (\$ 46 per ton) in the United States. For national gas the price range was \$ 0.78/MBTU to \$ 2.15/MBTU, and for fuel oil from \$ 1.95/MBTU (\$ 12.10/barrel) to \$ 2.03/MBTU (\$ 12.60/barrel).

Combined with the differences in fuel requirements observed above, this implies that the fuel cost difference between a wet and a dry plant could range from \$ 2.50 to \$ 16.50 per metric ton.

As far as electric power is concerned, the prices quoted ranged from 1.5 ¢/kWh to about 5 ¢/kWh in the United States and from 2.5 to 3.5 ¢/kWh in Sweden.

Table 7. Energy Consumption in Wet and Dry Process Plants

Plant nationality	Year of installation	Fuel consumption kWh/ton	Electricity consumption kWh/ton	Total energy consumption kWh/ton
Wet				
United States	1972[a]	1 775	145	1 920
United States	1960[a]	2 230-2 260	143	2 373-2 403
Sweden	1967	1 770[b]	129[a]	1 889
		1 689[a]		1 809[a]
Long dry				
United States	1970[a]	1 650	130-155	1 780-1 805
United States	1976[a]	1 520		1 650-1 675
United States	1965[a]	1 455	..	..
Suspension preheater				
United States	1976	1 160	185-210	1 345-1 370
United States	1976	1 100	175	1 275
United States	1973	970	..	..
United States	1974	970	..	..
Sweden	1969	940[b]	101[a]	1 041
Sweden (projected)	1979	930[b]	109[a]	1 039
Precalciner (projected)				
United States	1978	935	106	1 041

[a] Includes older part of plant.

[b] Latest kiln only.

As far as the empirical evidence on the relative labor saving is concerned, the picture is less clear. If figure 8 were taken at face value, it would indicate that labor costs are substantially lower in suspension preheater systems than in wet ones. However, there are simply too few observations to permit any conclusions. But in this case the interview results are unambiguous: there are no differences to speak of, given the scale of the installation. At most, there is a difference of one man per shift more in preheater systems (preheater attendant) than in wet systems. The cost difference would amount to only $ 10-0.20 per ton of cement.[1]

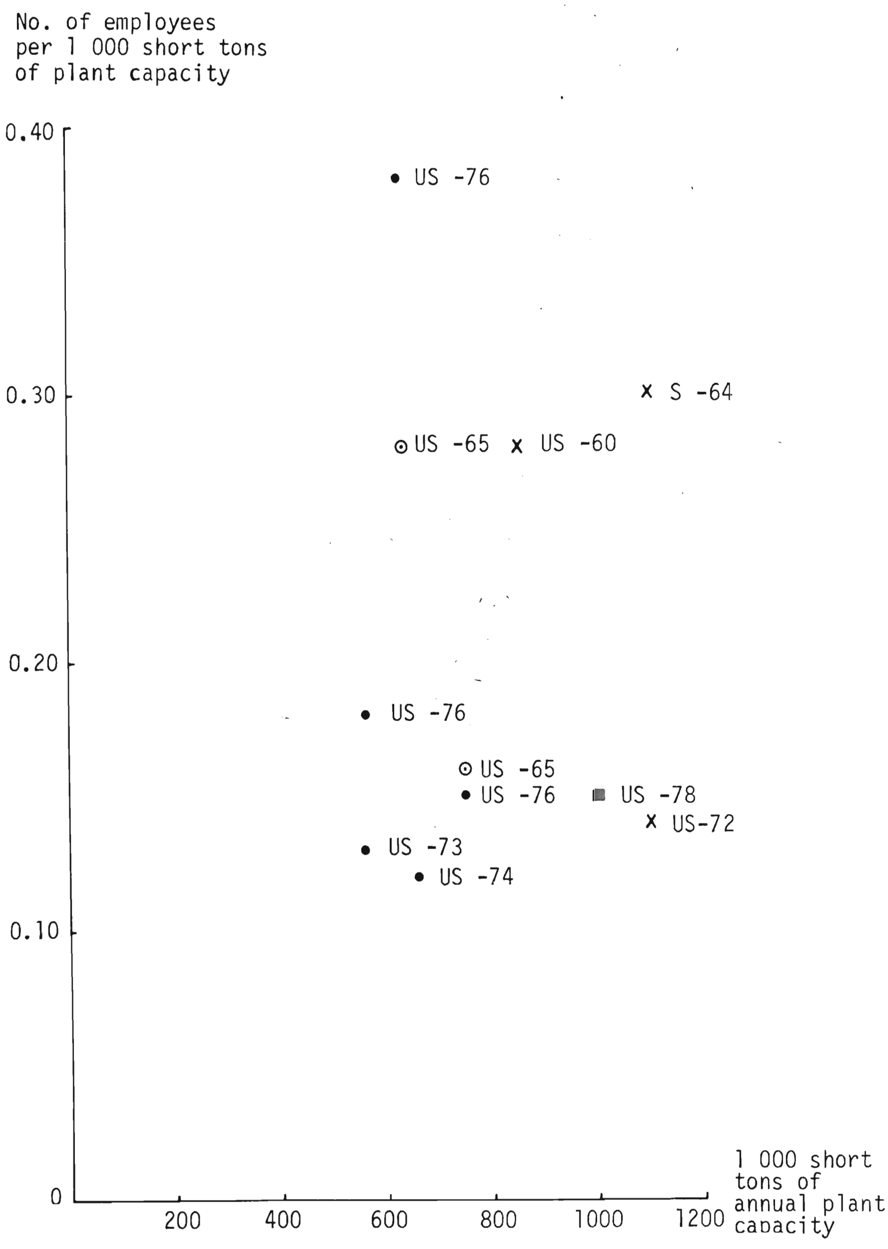
Turning to capital costs per ton of capacity, the evidence is much less clear. See figure 9. The figure has been constructed in the following way. The amount of the investment as reported by each company, has been divided by the (gross) additional capacity, yielding a raw figure on the capital cost per ton of annual capacity. Using information as to what items were included in the investment, it was estimated how much of the total investment for the standard plant given in table 8 was included, and the raw capital cost was adjusted accordingly. Then the adjusted figure was deflated or inflated by a price index to obtain 1974 prices. Unfortunately, no index of investment costs in the cement industry is available, so the United States Wholesale Price Index for industrial commodities was used. The fact that the estimated capital costs for late-year investments are found to be on the high side is probably an indication that some better price index must be found.

But even apart from the price index problem, it is difficult to make much sense of the data. It does not seem possible to say that one type of kiln has consistently higher or lower capital costs than another, nor is it clear even that capital costs decrease with scale. If anything, wet process kilns seem to have <u>lower</u> investment costs per ton than preheater systems. Invest-

---

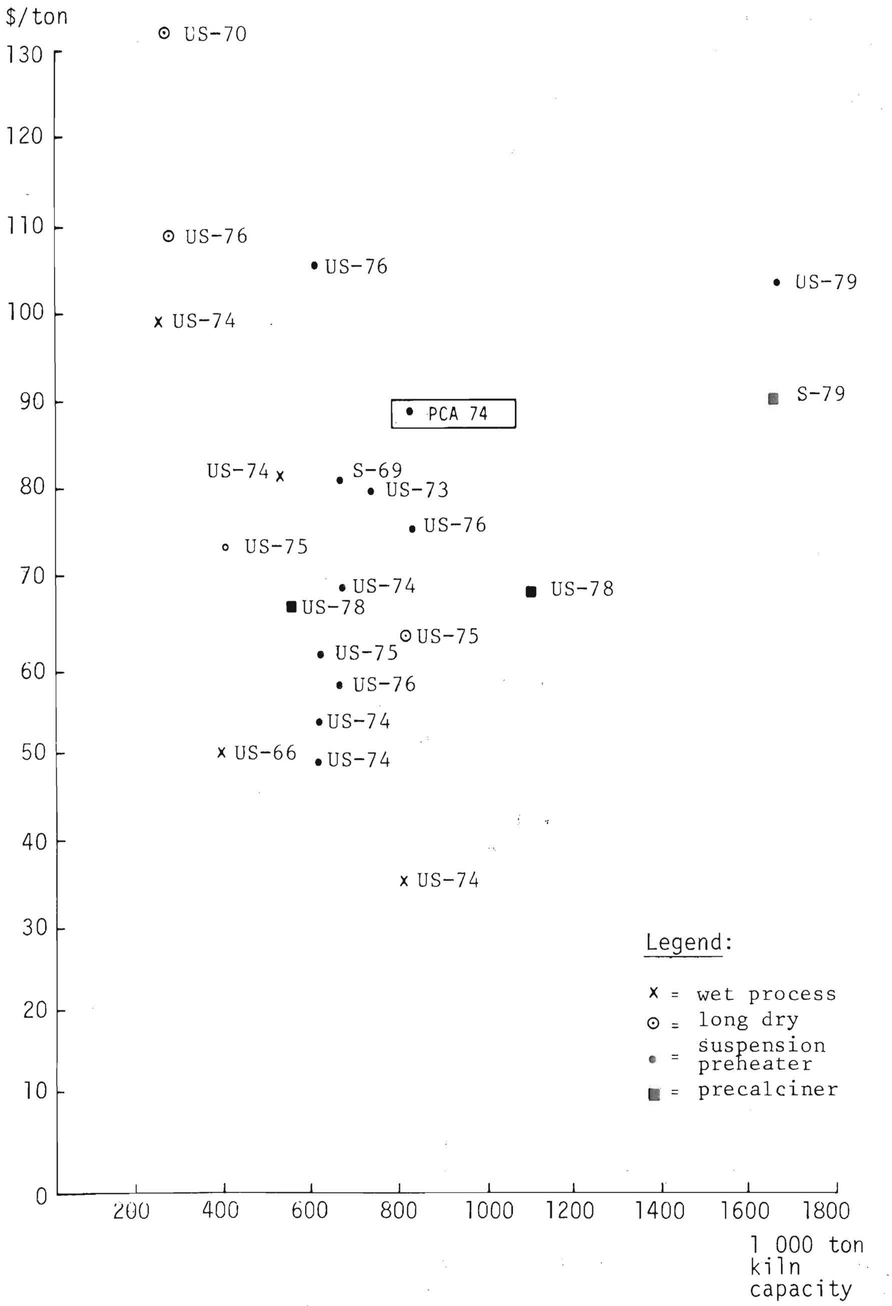
[1] Assuming three 8-hour shifts 330 days a year with a wage of $ 7.00/hour and an annual production capacity of .5 to 1 million tons.

Figure 8. Plant employment per ton of capacity



Note: Since it has not been possible to distribute employment on individual kilns, total plant employment is given, and the year given for each plant represents the average kiln age in the plant.

Figure 9. Investment cost per metric ton of capacity in cement plants
1974 prices



ment costs for precalciner systems seem to increase rather than decrease with scale, and the spread in investment costs for SP systems completed in 1976 is between \$ 52 and \$ 95 per short ton.

What conclusion can be drawn from these rather discouraging results concerning investment costs? Admittedly, the data are very crude, but it appears likely that no adjustment to standardize the data would be sufficient to obtain any observable pattern. There are apparently such large differences among plants that it is difficult to speak of a "standardized plant".

There are several reasons why investment costs vary widely among plants. Even though the standardized investment cost data in table 8 must be interpreted with great care, they at least indicate that the cost of installation is higher than the cost of the equipment. The installation involves various types of construction jobs -- supports for the kiln, buildings and roads, etc. -- the cost of which depends on local conditions (skill and efficiency of local contractors, ground conditions, etc.). In addition, the cost of the equipment varies substantially from one case to another. There are only a handful of cement equipment manufacturers in the world (one Danish, a few West German and Japanese; and two American companies which operate mainly on licenses from the other manufacturers) who compete in designing and selling whole systems. In order to obtain reference plants they are sometimes willing to offer extremely low prices coupled with substantial guarantees. And of course, prices are always set in negotiations between the cement firm and the equipment manufacturers.

The interview results indicate that opinions in the industry vary widely on whether wet or dry systems have lower investment costs. But it is clear that such statements usually reflect guesses rather than facts. A ong all the 14 interviews with cement firms in both the United States and Sweden concerning kilns or plants built in the last 10 years there was only one case in which a detailed comparison had been made of what a wet as opposed to a pre-

Table 8. Estimated Cost of a 2 200 Short Tons per Day Cement Plant Incorporating a Roller Mill and Suspension Preheater

Department	Equipment $ 1,000	Installation $ 1,000	Total $ 1,000	Total Percent	
Quarry equipment and amenities	4,000	300	4,000	7.4	
Limestone crushing	400	900	1,300	2.3	12.6
Limestone storage	500	1,150	1,650	2.9	
Raw grinding (roller mill)	2,250	5,200	7,450	12.9	
Additive and clay handling	600	1,400	2,000	3.5	
Blending	600	1,400	2,000	3.5	
Calcining	4,150	9,550	13,700	23.7	
Clinker grinding and gypsum handling	1,700	3,900	5,600	9.7	
Loadout and packing	600	1,400	2,000	3.5	
Electrical distribution and central process control	1,600	3,700	5,300	9.2	
Electric motors	1,200	2,750	3,950	6.8	
Land (640 acres)	1,000		1,000	1.7	
Storage facilities	1,000	3,000	4,000	6.9	
Land improvements	1,000		1,000	1.7	
Coal equipment	1,250	1,250	2,500	4.3	
Total	21,850	35,900	57,750	100.0	
Cost per ton of capacity			80		

Source: PCA Economic and Market Research Department.

heater dry installation would cost. In that particular case, the cost comparison came out 20 % lower for the suspension preheater system. But the investment covered only a capacity expansion, not an entire plant. If a whole plant had been considered, the relative cost difference probably would have been about half as large. In none of the interviews were capital cost considerations given as the main reason for choosing a particular process, and in no case was the investment cost difference between the chosen process and an alternative one deemed to be larger than 15 %.

This is not to say that investment cost differences are unimportant -- after all, even a 15 % saving on capital cost would amount to over $ 2 per ton of cement (i.e. about as much as the energy cost differential), if the previously calculated $ 14 per ton is a representative capital cost. But it is clear both that no careful comparison of investment costs was usually made and that fuel saving arguments were given in favor of preheater systems and raw material conditions in favor of wet systems.

To the extent that it is possible to draw any conclusion from this discussion at all, it would seem to be the following. Labor requirements play no role at all in choosing among the available technologies. Labor saving arises through increases in scale, regardless of which process is chosen. Even if it is true that capital cost considerations have not played any major role in choosing between alternative technologies in the United States, it is also true that U.S. cement installations in recent years have not been particularly large in comparison with European and Japanese plants. Instead, they have been in the size range where wet process kilns seem to have a comparative, even if not absolute, advantage. It is possible, therefore, that as plant and kiln scale continues to increase, capital cost considerations will become more important -- and labor cost differences as well. But up until now, energy savings seem to have provided the main argument for the preheater technology.

6. REASONS FOR THE DELAYED INTRODUCTION OF SUSPENSION PREHEATER KILNS

The previous discussion has indicated that the only argument for the suspension preheater technology which holds up under scrutiny is the fuel saving argument. Therefore, in order to justify continued investments in wet process plants, one would have to argue that the fuel saving argument was not applicable to the particular installation considered. There seem to be essentially four reasons why the fuel saving argument may not have been applicable in individual cases.

First at all, one factor which naturally affects the choice between wet and dry process is the moisture content of the raw materials. In our sample of plants, the moisture content varies from 1 % to over 20 %. The water content of the feed must be reduced to close to zero in any dry operation. In conventional raw grinding mills (so-called ball mills) there is enough heat generated in the grinding process, although no heat is added, to dry materials containing up to 7 % water.[1] Therefore, there seems to have been a long-standing rule of thumb in the U.S. cement industry that any material with higher than 7 % moisture content is unsuitable for the dry process.

However, a new type of grinding mill, so-called roller mills, was developed in West Germany, apparently in the early 1960's. This type of raw mill is used widely in Europe but was introduced in the United States only in 1973.[1] Roller mills use 5 to 15 % less electricity than ball mills, but they are also more amenable to combined grinding-drying than ball mills. By utilizing low-level heat in waste gases from the kiln or preheater it is possible to dry raw materials containing up to 15 % moisture.[2]

---

[1] U.S. Bureau of Mines, Minerals Yearbook, 1974, p. 298.

[2] Gordian Associates, op.cit., p. 14.

By installing additional heating equipment it is possible to dry raw materials with up to 18 % moisture content. The roller mill seems to have been developed precisely to increase the range of utilization of suspension preheater kilns.

At the present time it is not clear whether roller mills per se, require higher or lower investment costs than ball mills. But since they can grind feed of much larger size than ball mills, they may eliminate a secondary crusher which is usually required. Also, they operate at a much lower noise level than ball mills, (reducing the need for noise abatement equipment). Thus, overall it would appear that the capital cost of roller mills is probably lower than that of ball mills. The cost of equipment wear is reported to be about 60 % lower than for ball mills.[1]

The implications of this are that in cases where the moisture content exceeds 15 % there may have been reasons to choose the wet rather than the dry process. Even though it seems difficult to argue that the raw materials are wetter, on the average, in the Unites States than in, say Germany or Sweden, high overland transportation costs and absence of inland water transport facilities may have led to exploitation of wet materials which might not have been used at all in Europe. In the Swedish case, the geography has permitted all but one plant to be located near water, as was noted earlier.

Another problem which affects the choice between wet and conventional dry systems on one hand and suspension preheaters and precalciners on the other is the presence of certain substances in the raw materials which cause operational difficulties or affect the quality of the product negatively. The most important of these substances are alkalis (natrium and potassium). If cement containing alkalis is mixed with certain aggregates -- prevalent in the Southeastern United States but also in certain other areas, a chemical reaction occurs which causes the concrete to crack.

---

[1] U.S. Bureau of Mines, Minerals Yearbook, 1974, p. 298.

Therefore, the alkali content is regulated by law. The limit set in the United States is 0.6 %. However, even customers in areas without reactive aggregates often specify low alkali cement. Other countries also have restrictions on alkali content, although not as stringent. Efforts are currently being made in the Unites States to enforce the restrictions only when necessary.

But the presence of alkalis also creates problems in the manufacturing process itself. Since these are highly volatile substances, they will simply be blown out with the kiln exhaust gases in open systems such as wet or conventional dry kilns. But in suspension preheater or precalciner systems which are much more enclosed, alkali content builds up in the circulating air. If the alkali content of the raw material is low, or if there is just the right amount of sulfur in the raw material or fuel to balance the alkalis, there is no operational problem in the preheater. But if there is too much or too little sulfur, the preheater gets plugged up with sticky material which causes stoppages unless removed.

In order to prevent alkali buildup in suspension preheaters, a so-called by-pass has been developed which allows hot air with high concentrations of alkalis simply to escape from the system. This involves an additional investment cost and the loss of both energy and raw materials escaping with the hot air.

It is suggested by some sources [1] that at least some precalcining systems are capable of yielding low-alkali cement with difficult raw materials even with little or no by-pass. However, this is an issue which needs further investigation.

Given that high alkali content and presence of reactive substances do present difficulties in certain parts of the Unites States and much less in Sweden, the implication is that the alkali problem

---

[1] See e.g. Gordian Associates, op.cit., p.25.

explains at least some of the observed differences between the two countries in the attitude to the dry process.

The obvious question that arises is whether the alkali problems are unique to the United States or why these difficulties do not seem to have played the same role in other countries. But while it is true that the restrictions on alkali content are more stringent in the U.S. than elsewhere, it is difficult to believe that something as common in the crust of the earth as limestone could vary so much in quality or composition as to be unsuitable for a particular process on one continent but not on another. The following might be at least a partial explanation. Coal is the main fuel used in the cement industry in the United States, while in the 1950's and 1960's most European producers switched to oil. Due to the refining process, the sulfur content of fuel oil is held within very narrow margins, even for high-sulfur oils, which means that it is relatively easy to maintain a certain balance between sulfur and alkali in the cement manufacturing process. Coal, on the other hand usually contains much more sulfur, but above all, the variability of sulfur content is much greater.[1] This factor, in conjunction with the alkali restrictions in the U.S., provides a third reason for the relatively slow diffusion of suspension preheaters in the United States.

A fourth reason for the delay in introducing the suspension preheater technology, particularly in the Unites States, is the bad experiences that several companies had in their early efforts to introduce the new technology. In 1953, just three years after the first SP system was installed in Germany, the first preheater system was built in the United States. This was the fourth such system built in the world until then, which shows that U.S. producers were quick to respond to the new technology. The first SP kiln was followed in the next few years by twelve more. But the majority of these preheaters ran into several operational difficulties having to do with a lack of understanding of the sulfur-alkali

---

[1] Garrett, op.cit., pp. 273-277.

balance and similar problems. Consequently, many of these preheaters often clogged up, causing considerable downtime and thereby raising both capital and labor costs. About half of the thirteen original U.S. installations have now been shut down (some having been replaced by wet kilns!), and between 1955 and 1970 there were only two suspension preheater kilns sold in the United States, one of which has since been shut down.[1]

Ironically, therefore, part of the overcapacity in the 1960's was due to the installation of suspension preheaters, many of which did not function well. Both the overcapacity and the malfunctioning held back further investment in SP systems. And because of the operational difficulties, the belief became widespread that suspension preheaters were unsuitable to U.S. raw materials and operating requirements.

7. CONCLUSIONS

This study started out with the notion that a comparison between the United States and Sweden in the choice of cement production technology would be a simple illustration of substitution between energy and other factors of production in view of the differences in relative factor prices, especially relative energy prices. It was soon discovered, however, that the suspension preheater process can be regarded as theoretically superior to the wet process in almost all respects. The problem then became that of explaining why the rate of diffusion of the new process has differed among countries, particularly between Sweden and the United States. It was shown in a cost comparison of the wet and the dry process, based on national average data, that differences in relative factor prices must have been a major influence, and that the drastic price changes which took place between 1970 and 1975 probably

---

[1] Garrett, <u>op.cit.</u>, pp. 273-277.

constitute the major reason why investments in the wet technology have dwindled to zero.

However, it has also been shown that there are many factors which in actual practice reduce the theoretical cost differences quite drastically. The range of variation among plants in prices, raw materials, the market situation etc., is very large indeed. In addition, despite efforts to standardize for differences among plants in type and size of kiln, year of installation, etc., it proved very difficult to find any sensible patterns in the data other than with respect to energy.

The insights gained through this study relate to understanding the process of change within an industry, the forces which interact to generate this process, the interrelatedness of technical and market conditions, relative prices, rules of thumb, and the attitudes of decision makers in any investment decision, etc.

A final word about the future might be in order. Given the enormous spread between best practice and average practice plants in the United States - much larger than in Sweden e.g. - it appears highly probable that rising energy prices will lead to drastic structural changes within the industry. This process has already been analyzed for Sweden.[1] Interesting questions arise as to whether the U.S. cement industry will be able to effect the necessary changes, given the long history of experience with wet plants, the existing market structure, and the low profitability in recent years. There has been a number of cases recently of European firms buying up old U.S. plants in order to acquire market shares, then replacing them with new, larger equipment. This process is likely to continue unless prevented through government policy and is likely to yield a higher degree of both efficiency and concentration of ownership.

---

[1] See B. Carlsson, "Industrins energiförbrukning 1974-80" (Industrial Energy Consumption 1974-80), appendix 7 in IUI:s långtidsbedömning 1976. Bilagor (IUI's Long Term Survey 1976. Appendix Volume). IUI, Stockholm, 1976; pp. 277-287.

TECHNICAL CHANGE IN THE SWEDISH HYDRO POWER SECTOR 1900–1975*

Anders Grufman
IUI, Stockholm

A hydro power station has two functions from an energy point of view, on the one hand to make the potential energy of the water available for energy conversion (energy gathering) and on the other hand to perform the energy conversion from kinetic energy to electric power. The aim of this paper is to point out some main characteristics of the development of the energy conversion stage and to give some quantitative measures of the energy saving technical change in this stage.

Even though the energy gathering stage and the energy conversion stage are constructed interdependently we shall, however, at first discuss the energy gathering stage. The easiest way to do this is to start with the physical relationship that expresses the relation between energy (E) quantity of water (m-kilogrammes) and head (h-meters)[1] in a waterborne energy resource.

$$E = m \cdot h \cdot g \text{ (Ws = Wattseconds)} \qquad (1)$$

where g = acceleration of gravity force 9.81 (m/s^2).

[1] Drop.

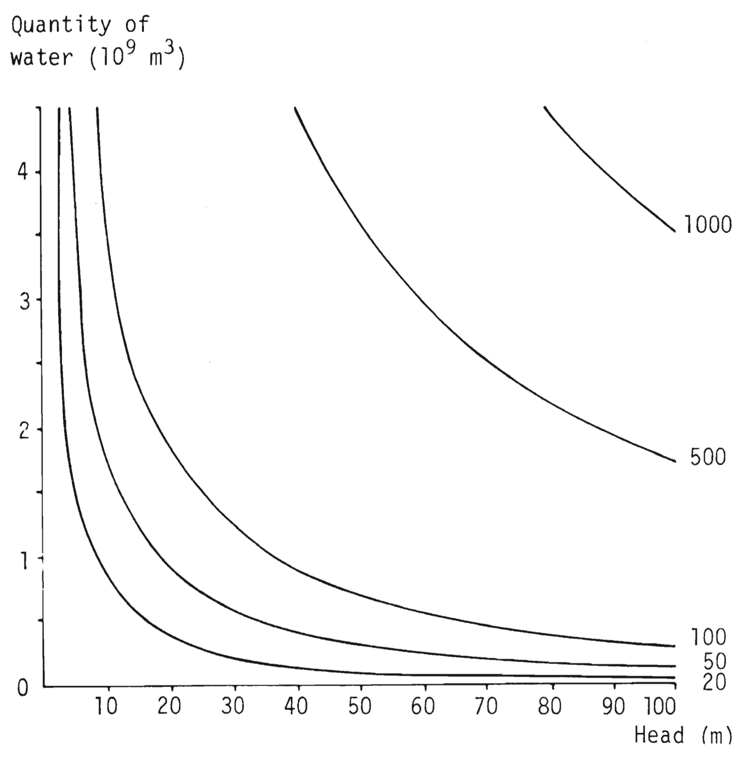
* This paper derives from a larger project on technical change in the Swedish energy conversion sector that the author has undertaken at the IUI. Teknisk utveckling och produktivitet i energiomvandlingssektorn (Technical change and productivity in the energy conversion sector), IUI, Stockholm 1978.

I would like to thank Jim Albrecht, Lennart Hjalmarsson and especially Bo Carlsson and Leif Jansson for advice and comments on earlier drafts.

This relation can also be seen in Diagram 1. The isoquants unite different combinations of m and h, which give the same energy (E). We can think of this relation as an equation expressing the energy quantity in a shallow lake with m-kilograms of water h-meters above a lake. In principle there exists full interchangeability between increasing the quantities of water (m) on the one hand, and increasing head (h) on the other in order to gather a certain quantity of natural energy resources. Natural conditions are very important in determining whether a certain quantity of energy is going to be produced in a power station, say with large quantities of water and a low head.

When blasting and constructing techniques were undeveloped, the dimensions of a power station were more restricted by natural con-

Diagram 1. Isoquants (in Gwh=10^6kWh) referring to different combinations of head and quantities of water



ditions than is the case today. Economic possibilities now exist for blasting long tunnels and building large water accumulating systems.

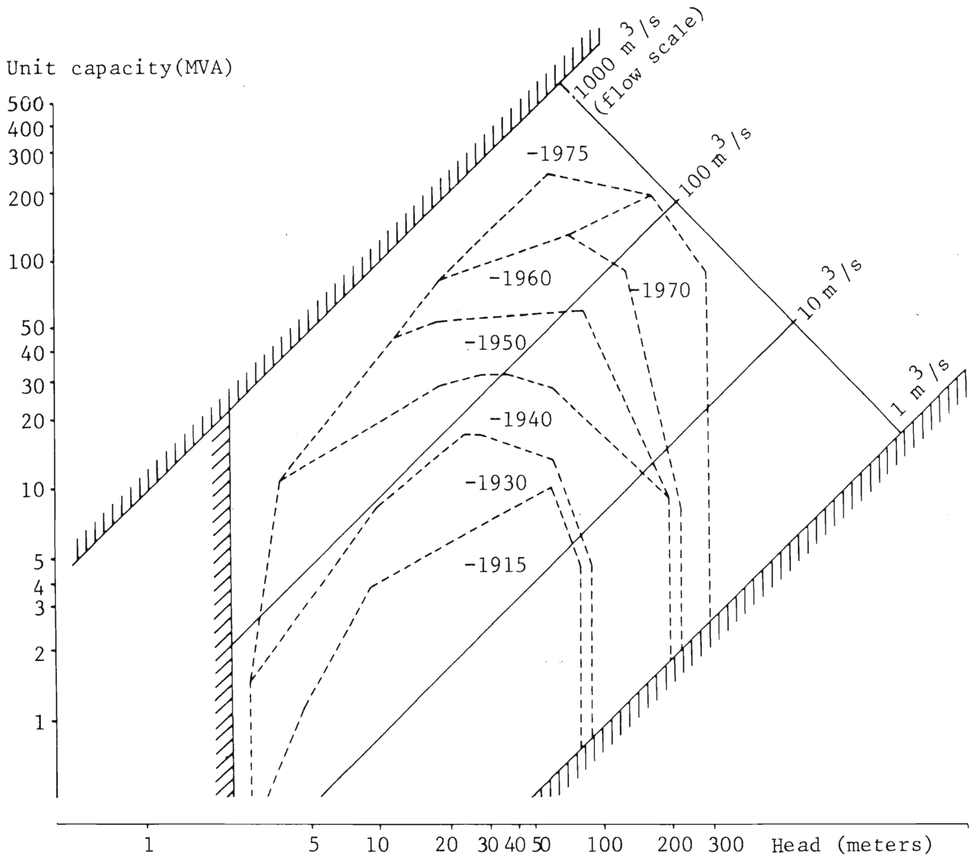
The energy conversion technique chosen for a power station thus depends greatly upon the "energy gathering" technique, in combination with the natural conditions. Therefore we cannot speak of a "best practice" plant for energy conversion in the traditional sense. It is usually assumed that the best practice plant is optimal in scale and technique with respect to present price and technology expectations. Hydro power energy conversion is furthermore restricted as regards inputs into the process. Input to the energy conversion stage does not consist of the homogeneous input "hydro energy" but rather of "hydro energy at a certain pressure". Since energy in different forms requires different construction of the energy converting equipment, different heads (h) will demand different turbine and alternator designs - with a given state of technology and capacity. The best practice energy conversion plant will thus be the plant that with given capacity and head demands the lowest amount of resources, mainly in the form of energy- and capital inputs because modern hydro-electric power plants are mostly unmanned.

TECHNICAL CHANGE IN THE ENERGY-CONVERSION STAGE

Since every new combination of capacity and head (and therefore also quantity of water) for a hydro-power unit represents a new mode or technique of production, one aspect of the technical change in the energy-conversion stage is therefore how the "frontier" of these combinations has moved over time. The other aspect is how input of resources has varied over time, given these combinations.

To start with we shall study how this technical frontier has developed. (See Diagram 2.) We have used data from 841 units installed between 1900 and 1974. Their total capacity is

Diagram 2. Limits to flow, head and capacity-characteristics for hydro-power units installed 1900-1975



14 823 MVA[1], which includes most of total installed capacity during this period. In Diagram 2, which is double logarithmic, head (in meters) is along the x-axis and installed unit capacity along the y-axis. Because E, h and m are multiplicatively related, points with the same flow will form straight lines in this diagram (iso-flow lines). The scale on the right refers to the flow (m^3/s). To give an impression of the possible space of

[1] MVA (Megavolt ampères) is equivalent to MW corrected for a coefficient expressing the offset in phase between current (in ampères) and voltage. This offset is measured as an angle ϕ. The correction coefficient becomes $\cos(\phi)$ and usually takes values around 0.9.

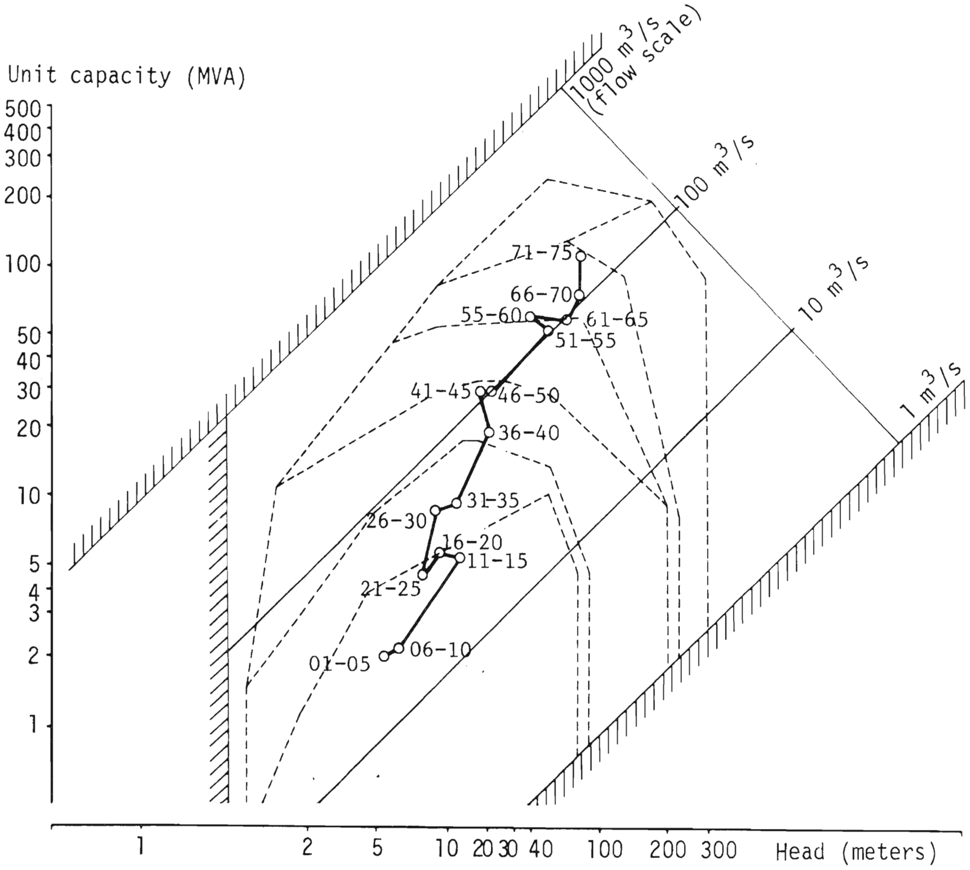
existence for units we have marked the present space of existence for power plants with limitation lines. (A power plant can consist of one or more units.) Thus, e.g., the lowest head was 2.5 m and the largets and the lowest flows were 1000 m^3/s and 1 m^3/s, respectively. These "limits" should, however, not be considered as absolute in the sense that they cannot be crossed. Within this region we have marked the combinations of head, waterflows and unit scale installed up to a certain period. The material has been subdivided into the following periods: 1900-1915, 1916-1930, 1931-1940, 1941-1950, 1951-1960, 1961-1970 and 1971-1975.

As can be seen in Diagram 2, the most pronounced characteristic of the development is the increase of the unit scale. This can be seen as an upward shift of the maximum attainable Megawattage for each period. The vertical distance between the upper point of each period is roughly the same for all the chosen periods. This indicates that the rate of growth in the maximum scale has been approximately constant over a long period, even though it has been slightly quicker during the fifties. The average of these vertical distances implies a near doubling of maximum scale during each period. Analogously an increase in the maximum flows can be seen as a shift of the limitational lines perpendicularly with respect to the iso-flow lines. The relative growth of the maximal flow stops almost completely already after the 1940's.

Changes in maximum head are seen as a shift in the rightmost limitation lines along the x-axis. The quickest growth of head occurred between 1930 and 1940. Already during the 1930's high heads were used for hydro power production. This picture of the development can be complemented by studying the characteristics of the average capacity installed. For the sake of clarity these can be seen in a new diagram (Diagram 3).

Every circle in the diagram represents the average unit characteristics (scale, head, flow) during each five year period between 1900-1975. (Every five year period consists of 25 to 114 observations.) The averages have been calculated by weighting with the unit capacities.

Diagram 3. Average flow, head and capacity of units installed 1900-1975



To begin with we can see that average installed head has increased from 10-20 meters in the beginning of the century to 80-90 meters during the seventies. Average unit rating has increased from 2-3 MW to 100-150 MW during the same period. From a capital productivity point of view this means that during this period the volume of water handling per unit of production has decreased substantially.

If we study the pattern of development from period to period, we can, however, note some irregularities. Between 1916-25 no increase in neither the unit rating nor the head does occur. One explanation may be the large increase in construction costs which ocurred in connection with the first world war. Therefore, the construction of hydro power stations which required more building work was avoided. A similar tendency can be seen during the second world war. During the period 1925-1940 construction costs decreased, which is also reflected in a substantial increase towards larger waterflows and higher heads.

During the period 1950-1955 capacity expansion was almost solely achieved by higher heads, that is, parallell to the iso-flow lines. After 1955 the pattern is more irregular. During 1955-1960 a worsening of the head conditions occurs. Between 1961 and 1965 erection of capacity with favourable heads takes place but with lower average natural energy resources. During the following 10-year period, 1966-1975, an increase in the average rating of units occurs, however, without the usual simultaneous improvement in head conditions.

This could be related to a strong change in the relative cost of expanding energy capacity from increasing heads to increasing quantities of water. Such a relative change in costs could occur if topographic conditions are altered in such a manner that it is less costly to expand water accumulation capacity,(e.g. by connecting adjoining precipitation regions) than to blast long mountain tunnels to attain heads. (After 1960 capacity expansion took place mainly in the northern Norrland region which has more favourable topographic and hydrological conditions than southern Sweden in this sense.) The above mentioned shift in relative costs can be viewed as a change in price of water relative to head. As can be seen in Diagram 4 a drastical increase in average dam volume (weighted by size in MW) took place after 1955. An increase in dam volume can be seen as a rough proxy for the size of interconnected water systems. Furthermore, it can be seen in Diagram 5 how average tunnel length per head meter has developed. Until the end of the 1960's this ratio increased relatively slowly compared to the

Diagram 4. Average dam-capacity per MW installed 1941-74

Index for period before 1941 = 100

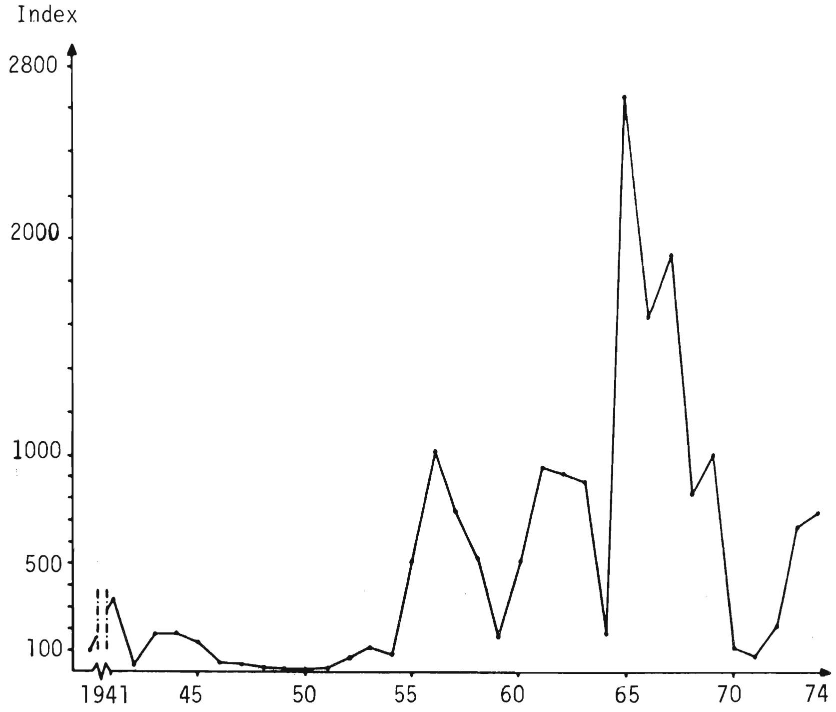
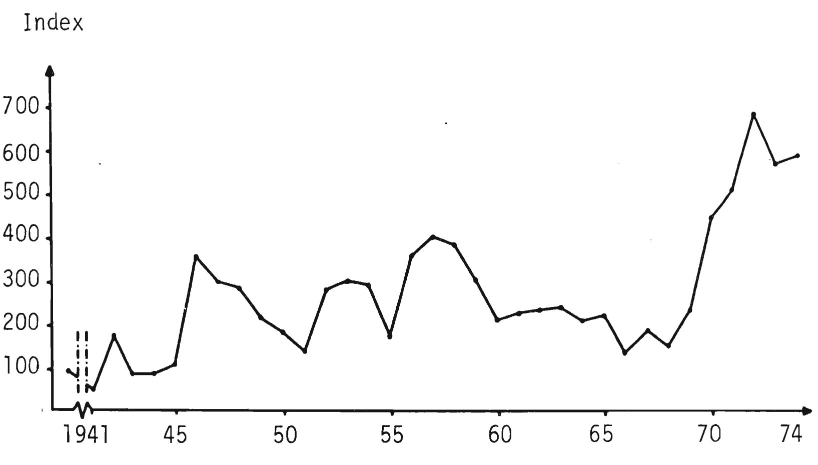


Diagram 5. Tunnel-length per head-meter and MW installed 1941-74

Index for period before 1941 = 100



period 1969-1975, when it grew from approximately 200 to a high of 700. In order to achieve the 1972 average head of 85 meters (see Diagram 3) one had to blast 6.5 km of tunnels. (All other waterways, canals etc. not included.) This can be compared with 1.8 km of mountain tunnels to achieve an average head of 65 meters during the 50ies.

This very rapid change (worsening) of the conditions for expansion is probably one of the main reasons why expansion of hydro power in Sweden has almost halted during the 70ies.

ENERGY PRODUCTIVITY AND TECHNICAL CHANGE

In the following section we shall give an account of an attempt to measure and quantify the energy-saving technical change in the hydro-power sector. The data refer to 263 plants built between 1900 and 1974. Efficiency measures refer to cross-section data in 1974. The long life of hydro-power units has made it possible to make estimations for plants of high age. (Plants built before 1900 are still used for commercial production.)

Successive repairs and improvements, however, have increased efficiency in the oldest plants in such a manner that their efficiency in 1974 does not reflect their efficiency at the date of construction. There is, though, no good way to know how much this error affects our estimates. The technique of measuring efficiency in hydro-power stations is much younger than the technique of producing hydro power. Some complementary investigations seem to suggest that even if the cross section analysis biases our estimates of the speed of growth in energy efficiency downwards this bias is of minor importance due to a strong "embodiedness" of the technique for each vintage.

Specification of a partial engineering production function

We shall view technical progress as the shift over time of the (ex-ante) function expressing the relation between input and output of energy. The energy saving technical progress analysis will be performed by studying energy conditions solely at the unit level. In order to refer to the measured increase in energy efficiency (that is, the ratio between input and output of energy) as technical progress, we have to make the assumption that the possibility of substitution between energy and capital at the unit level is small. This implies that substitution between energy and capital at the plant level takes place by choosing the number of units in a station. This type of substitution is possible due to the rather surprising fact (which shall be demonstrated later in this paper) that gross capital requirements of the energy conversion equipment at the plant levels decrease with the number of units given the capacity of the plant. An increase in the number of units, given plant capacity, will, however, reduce plant energy efficiency due to scale effects at the unit level. Therefore there will be a tradeoff at the plant level between energy and capital, but not at the unit level.

The general production function relation for a hydro power unit is assumed to be expressed in the following way

$$f(x, y, k_1, \ldots, k_n) = 0, \qquad (2)$$

where x = input in the form of natural energy (motive power)
y = output in the form of electric energy (power)
$k_1, \ldots, k_n$ are design parameters.

The function could most adequately be described as an "engineering production function", because it includes the effect on energy productivity of among other factors, the design of water systems and type of turbines.

In the following these variables will be used:

C = capital (investment)

E = energy of a waterborne energy resource in Wattseconds

g = acceleration of gravity force = 9.81 (meter/sec^2)

h = gross head in meters

l_v = length of tunnels in meters

l_v/h = length of tunnels l_v relative head h

$\ln P_e$ = scale parameter expressed in logarithmic form in order to take account of decreasing $\Delta\eta/\eta$ with increasing unit scale. (Decreasing elasticity of scale.)

m = quantity of water in kilogrammes

$\dot{m}$ = flow of water in kilogrammes per second

P_e = Unit capacity in MW or MVA

R = Dummy variable taking value 1 if turbine can be regulated (Kaplan) and 0 if it cannot (Francis). Since we do not know if the installed turbines are of the Kaplan or of the Francis type, we have assumed that if the unit was installed after 1935 (year of introduction of Kaplan turbines) and the head is lower than 15 meters, then the turbine is of the Kaplan type. When this proxy variable was compared with the true value for a smaller sample of units, however, we achieved a correlation of only 0.28.

t = unit age

β_1 = capital coefficient

β_2 = head coefficient

γ = shift coefficient

* expresses conditions during maximum production in a plant.

The statistical model

The energy-loss function used as a starting point for the statistical estimations is

$$\frac{P_e}{\dot{m}^{*}h^{*}g} = \eta(l_v/h,\ \ln P_e,\ t,\ R)^{1} \qquad (3)$$

* signifies conditions under maximum production. The left-hand side term is the actually observed energy-efficiency of the plant under maximum-production conditions. After differentiation the relation between the loss function η and changes in the relative length of tunnels variable (l_v/h), unit scale variable ($\ln P_e$), unit age variable (t) and type of turbine variable (R) can be expressed:

$$\eta_j = \eta_{oj} + d\eta_s = \eta_{oj} + \frac{\partial\eta_j}{\partial(l_v/h)}\,d(l_v/h) + \frac{\partial\eta_j}{\partial(\ln P_e)}\,d(\ln P_e) + \frac{\partial\eta_j}{\partial t}\,dt \qquad (4)$$

where j attains different values for Francis and Kaplan turbines, respectively.

For statistical estimation of the partial derivatives of this equation we write the statistical model:

$$\eta_i = \alpha + \beta_1(l_v/h)_i + \beta_2(\ln P_e)_i + \beta_3 t_i + \beta_4 R_i + u_i, \qquad (5)$$

where u_i is an error term with $F(u_i) = 0$ and $E(u_i^2) = \sigma^2$.

Thus we have taken care of the two different techniques (R = 1 and 0) with a dummy variable and by assuming equal coefficients for the other independent variables. In this model technical change is included as a linear function of time. Since we cannot assume

[1] To go from the energy relation (1) to the momentaneous power relation (3) one takes the time derivative of (1) assuming constant head (h) that is $P_e = \frac{dE}{dt} \approx \frac{dm}{dt}\cdot h\cdot g\cdot\eta$, where dm/dt is massflow per unit of time, that is kg/s, (density of water is assumed to be 1 ton/m^3) and multiplied by the loss factor η.

a linear relationship over a longer period of time we have, besides the above regression equation, also estimated an equation in which every vintage has its own dummy variable (48 vintages between 1900 and 1974).

The results of estimations

The estimates of the coefficients according to equation (5) can be seen in Table 1.

Table 1. Energy productivity in the hydro-power sector 1900-1974
Explanatory variables and regression results

Intercept	Regression coefficients: Unit scale ($\ln P_e$)	Unit age (t)	Relative length of tunnels (l_v/h)	Type of turbine (R)	R^2	Degrees of freedom
0.8013	17.5×10^{-3}*** (11.6)	-10.5×10^{-4}*** (-7.1)	-24.2×10^{-6} (-0.3)	-20.7×10^{-5} (-0.04)	0.60	258

Note: t-value within parenthesis. *** = significance at the 1% level.

The coefficients for unit scale ($\ln P_e$) and unit age (t) are both significant and of the correct sign. The coefficient for relative length of tunnels (l_v/h) has the proper sign but does not significantly differ from 0. It is also doubtful whether the coefficient is of the correct magnitude. Its size implies that head losses in tunnels are 0.024 meters per kilometer of tunnel, whereas direct measurements of the losses show that they should lie around 0.5 meters per mountain tunnel kilometer[1]. The coefficient of type of

[1] Elfman, S., Vattenledande bergtunnlar vid kraftverk. Statens Vattenfallsverk. Stockholm 1975. Technical report.
In a mountain tunnel, friction losses are a function of the velocity of waterflow. With a given flow (in m^3/s) the velocity of flow will be a function of the cross sectional area. Since cost per tunnel km increases with increasing cross section one is usually forced to make a trade-off between tunnel cross section and energy losses, or generally speaking between capital and energy.

Cont.

turbine (R) is also insignificant, but not therefore, uninteresting. It implies an aspect of the relation between natural conditions and energy productivity, namely that energy productivity under stationary conditions is not importantly altered if plants are built to make use of high or low heads. The value of this coefficient could, however, depend largely upon the chosen proxy. The scale coefficient (lnP_e) implies that with otherwise equal (natural) conditions a doubling of unit scale leads to an increase in energy productivity with 1.3 percentage units. The unit age coefficient (t) shows that energy productivity, on the average, has increased with 1 percentage unit every 10 year.

As an example we can calculate with these values that a plant that was built in the beginning of the thirties with a unit size of 6 MW without tunnels should have had an energy productivity of approximately 0.79, while a unit built in 1967 of 220 MW with 5 km tunnels should on the average have an energy productivity of 0.89, both being operated at full capacity production.

We return to the matter of the coefficient for relative length of tunnels (l_v/h). Our estimate has a 20 times lower value than would be expected from physical measurements of tunnel losses. The reason we have this error is probably that it is difficult to separate the effects of unit scale (lnP_e) and unit age (t) from the effects of relative length of tunnels (l_v/h) in the regres-

---

Footnote 1 cont.

(There is, besides the possibility of increasing cross-sections, also the possibility of reducing flow losses by improving the surface conditions of the tunnel.) The point one chooses depends largely upon the natural rock-conditions (hardness, crackformations, etc.). These factors imply that losses per unit of tunnel length will vary between tunnels. The spread in friction does, however, not seem to be very significant. The average flow velocities at maximum production conditions lie around 1-1.5 m/s. The corresponding friction losses are on the average 0.5 m/km.

sions, because during a relatively short period (approximately since the middle of the 1960's) there has been simultaneous increases in both tunnel lengths and unit scale. Therefore the unit scale ($\ln P_e$) and age (t) variables have "explained" a part of the energy productivity decrease which undoubtedly has taken place as a result of increased tunnel lengths. We should for this reason assume that the unit scale ($\ln P_e$) and unit age (t) coefficients have been underestimated. One way to reduce the effect of this multicollinearity problem is to specify a new dependent variable $\hat{\eta}$, which is the observed energy efficiency at maximum production plus the expected value of the waterway losses that is

$$\hat{\eta} = \eta + \frac{l_v}{h} \cdot 0.5 \cdot 10^{-3} . \qquad (6)$$

Due to this we now have only unit scale ($\ln P_e$), unit age (t) and type of turbine (R) as independent variables. The results of this new regression can be seen in Table 2.

Table 2. Energy productivity in the hydro-power sector 1900-1974
Explanatory variables and regression results.
$\hat{\eta}$ = dependent variable

	Regression coefficients				
Intercept	Unit scale ($\ln P_e$)	Unit age (t)	Type of turbine (R)	R^2	Degrees of freedom
0.8133	19.2×10^{-3}*** (12.0)	-12.5×10^{-4}*** (-8.0)	-75.7×10^{-4}* (1.5)	0.64	259

Note: t-value within parenthesis. *** and * = significances at the 1% and 10% level, respectively.

As we see in Table 2 the coefficients of unit scale ($\ln P_e$) and unit age (t) increase somewhat with this operation as could be expected.

The results of the regressions with dummy variables for each vintage group can be seen in Table 3. The results also are suitable for graphic representation. We have plotted the intercept terms for the 47 vintage groups in Diagram 6. Not unexpectedly the linear trend through these intercept terms has the same slope as the OLS regression coefficients (see Table 2). We have plotted this trend in Diagram 6. Perpendicularly from this trend line we have drawn the lines which show the effect of (increasing) scale upon energy productivity. Clearly the increase in unit scale has meant roughly as much for the energy productivity development as has the general trend of the energy saving technical change.

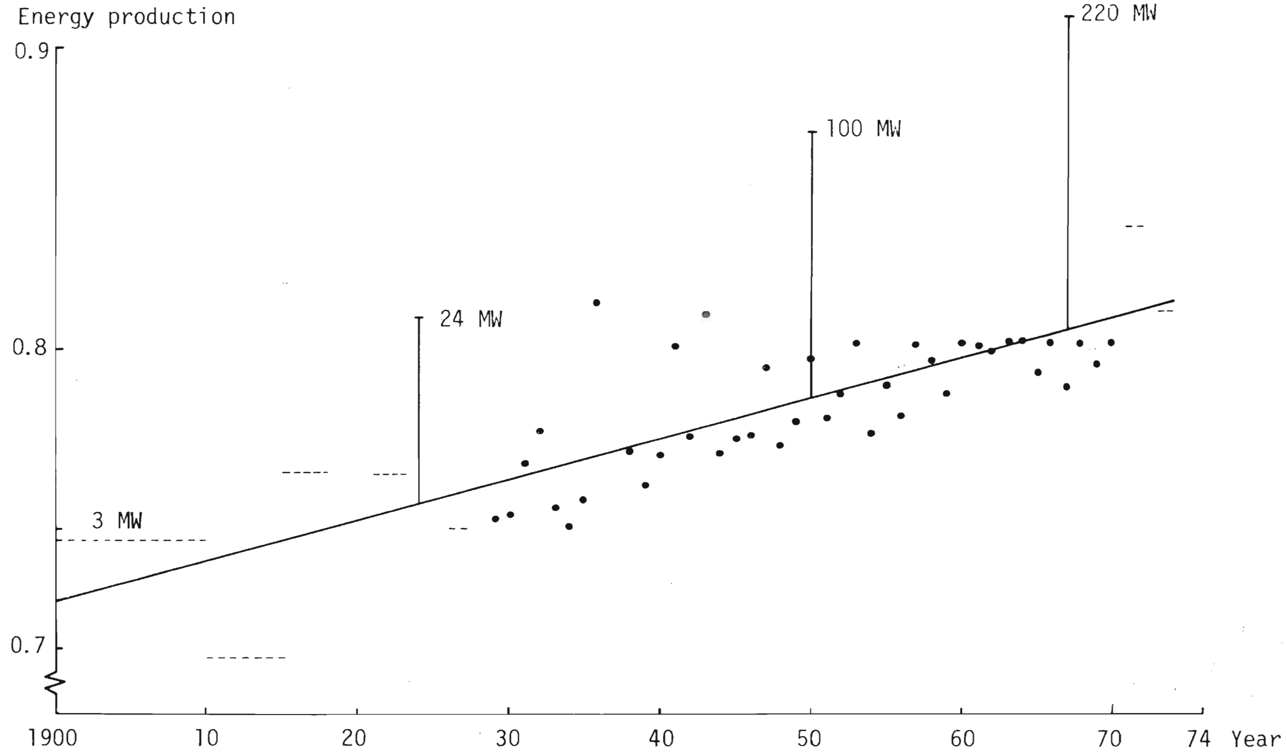
Table 3. Energy productivity in the hydro-power sector 1900-1974
Explanatory variables and regression results. Statistical model with individual term for each vintage
$\hat{\eta}$ = dependent variable

Regression coefficients			
Unit scale ($\ln P_e$)	Type of turbine (R)	R^2	Degrees of freedom
19.4×10^{-3}*** (10.8)	-105.2×10^{-4}** (-1.9)	0.43	213

Note: t-value within parenthesis. Intercept term, see Diagram 4.
*** and ** = significance at the 1% and 5% level, respectively.

Diagram 6. Energy-productivity, unit scale and technical change 1900-1974

Every dot or line represents an intercept



AN ASSESSMENT OF THE ENERGY-SAVING TECHNICAL CHANGE IN THE HYDRO-ELECTRIC POWER SECTOR

We shall ultimately try to calculate how much this energy saving development could be worth.

Suppose that we build a power plant with a capacity of 200 MW. We can build it with one unit and with an average utilization of 5 000 hours per year. Yearly production will then be 1 TWh (= one terawatthour which is equal to 10^9 kilowatthours), or we build it with two units of 100 MW each. According to our findings the two smaller units will produce with a 1.3 percentage units lower efficiency. Since energy efficiency in the first case will be on the average 0.90 (1970-vintage) the relative decrease in energy productivity will be 1.4 %. This means that for a given amount of supplied energy the two-unit station will produce 14 GWh (gigawatthours = 10^6 kilowatthours) less per year. The yearly worth of this production is in 1968 prices (0.032 Skr/kWh high voltage price excluding distribution costs) approximately 0.45 million Skr. Calculated with an average length of life of 30 years at an interest rate of 8 % this represents a capitalized value of 5.0 million Skr. This is to be compared with the average investment of approximately 10 million Skr in turbines and alternators in the one unit case. The investment in the two unit case thus has to be approximately 50 % lower in costs in order to justify the use of two units.

How do capital requirements vary with scale and head in the conversion stage? Some preliminary results referring to investments in the energy conversion stage suggest that the adequate specification of the relation between capacity and capital (unit capacity cost) is

$$P_e = C^{\beta_1} h^{\beta_2} e^{\gamma t} A, \qquad (7)$$

i.e., a Cobb-Douglas type of function where A is the intercept, the β's are the "marginal production elasticities" of capital and head and γ is a shift factor which expresses the rate of shift in this investment relation, due to, among other factors, inflation and capital saving technical change. We have estimated this relation using data referring only to turbines (49 observations) and to plants (25 observations). The turbines estimation gave the following coefficients (see Table 4). Note that capital is measured as investment in current prices, which leads (if prices have increased substantially on this type of equipment) to a negative sign on the shift coefficient γ.

Table 4. Turbine investment function. Estimated coefficients Turbines installed 1934-1975

Regression coefficients				
Capital coefficient (β_1)	Head coefficient (β_2)	Shift coefficient (γ)	R^2	Degrees of freedom
0.75*** (10.9)	0.54*** (10.6)	-0.008 (-1.43)	0.99	44

Note: t-values within parenthesis. *** = significance at the 1% level.

The unit regression is similar, but investment in this case refers to total investment in machinery per unit in the plant and not only to the energy conversion equipment.

Table 5. Machinery per unit investment function. Estimated coefficients. Units installed 1950-1974

Regression coefficients				
Capital coefficient (β_1)	Head coefficient (β_2)	Shift coefficient (γ)	R^2	Degrees of freedom
0.52** (2.3)	0.44*** (5.1)	-0.0159 (-0.9)	0.99	21

Note: t-value within parenthesis. *** and ** = significance at the 1% and 5% level.

Both regressions suggest that with given head and total capacity, capital requirements decrease with the amount of units installed, since the capital coefficients (β_1) are less than 1. Note, however, that in this step we have not considered the fact that capital requirements increase if building capital is included, since machinery takes more space if divided into more units. This is, however, important only in cases where machine rooms have to be blasted in the mountain.

Returning to our example, if we use two units instead of one, investments in machinery will decline (head is constant). If we use the results from Tables 4 and 5 investment requirement would decrease by between 50 and 25 % depending upon which of the capital coefficients (β_1) is considered the most reliable estimate. These investment reductions imply, with the figures given in our example, that investment could be reduced by 2.5-5.0 million Skr by using two units. If this is compared with the capitalized value of energy savings of 5.0 million Skr we arrive at a situation in which the choice very much depends upon the price assumptions we have made. The example, however, shows the great importance of energy productivity increase in the hydro-power sector. It also shows that energy productivity has played an important role in the process of increasing unit scale of production.

REFERENCES

Carlsson, B., 1977, Economies of Scale and Technological Change: of Blast Furnace Technology. Nijenrode Studies in Economics. Vol. 2 1977.

Chenery, H.B., 1948, Engineering Production Functions. Quarterly Journal of Economics 1948, pp. 507-531.

— 1953, Process and Production Functions from Engineering Data in Leontief et al., Studies in the structure of the American economy. Oxford University Press. New York.

Grufman, A., 1978, Teknisk utveckling och produktivitet i energiomvandlingssektorn (Technical change and productivity in the Swedish energy conversion sector). IUI, Stockholm.

Hjalmarsson, L., 1973, Optimal Structural Change and Related Concepts. Swedish Journal of Economics 75 (2), pp. 176-192.

Huettner, D., 1974, Plant Size, Technological Change, and Investment Requirements. A Dynamic Framework for the Long-run Average Cost Curve. Praeger Publishers. New York.

Marsden, S., Pingry, D. & Whinston, A., 1974, Engineering Foundations of Production Functions. Journal of Economic Theory, No. 9 1974, pp. 124-140.

Nabseth, L. & Ray, G. (ed.), 1974, The Diffusion of New Industrial Processes. An International Study. Cambridge University Press. London.

Nadiri, I., 1970, Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey. Journal of Economic Literature, Vol. VIII, December 1970.

Sveriges utbyggda vattenkraft 1930; Specialundersökning av Kommerskollegium. SOS, Stockholm 1931.

Vattenkraft och miljö; SOU 1976:28. Stockholm.

TECHNICAL PROGRESS AND STRUCTURAL EFFICIENCY OF SWEDISH DAIRY PLANTS*

Finn R Førsund and Lennart Hjalmarsson
University of Oslo IUI, Stockholm

SUMMARY

Technical change is estimated within a frontier production function allowing neutrally variable scale elasticity. To facilitate an analysis of structural change an average function is also estimated.

The results give little support for a hypothesis of neutral technical progress but rather a pattern of technical progress due to labour saving technical change increasing marginal productivity of capital relative to labour. The comparison between best-practice and average practice estimates also reveals an increased difference between best-practice and average practice techniques.

Numerical measures of the distance between best-practice and average practice are computed. Moreover, Salter's measures of bias and technical advance are also generalized and computed.

---

* Paper presented on the international colloquium on Capital in the Production Function at Paris X - Nanterre Novembre 18-20,1976.

We wish to thank H.J. Bakke and R. Teige for valuable programming and computational assistance.

Earlier versions have been presented at the Nordic Meeting of economists, Marstrand, May 1976, and at the Econometric Society European Meeting in Helsinki, August 1976.

The present version has benefitted from comments received at the Nanterre colloquium.

1. INTRODUCTION

The purpose of this paper is to put forward some results from a research project in industrial structure and structural change (in the spirit of Johansen [1972]) based on time series data from Swedish dairy plants.

To bring out the structure of the industry both average, (AP), and bets-practice, (BP), production functions are estimated. The use of combined cross-section - time-series data allows investigating technical change both for the AP and the BP functions. (In the literature (as far as we know) no explicit attempt to estimate technical change within a frontier function has been made (see e.g. Aigner & Chu [1968], Carlsson [1972], Timmer [1971]). The functional form chosen is the homothetic function, which permits the study of scale economies. (The programming estimation method of Aigner & Chu [1968] is generalized to handle this specification.)

2. ESTIMATION OF BEST PRACTICE FUNCTIONS

When estimating frontier functions three general approaches are found in the literature (see Johansen [1972] chapter 8 for a critical evaluation of some of the approaches): i) utilizing the whole sample, but restricting the observed points in the output-input space to be on or below the frontier, ii) eliminating "inefficient" observations and estimating an "average" frontier function from the subset of efficient points, iii) allowing some observations to be above the frontier either by eliminating a certain percentage of the most efficient observations (fitting a "probabilistic" frontier à la Timmer [1971]) or specify both an efficiency distribution proper and pure random variation of efficiency (see Meeusen & van den Broeck [1976]).

We will here utilize approach i) and generalize the programming method in Aigner & Chu [1968] to allow for neutrally variable returns to scale.

The best-practice production function is pre-specified to be a homothetic function of the general form

$$G(x) = g(v,t), \tag{1}$$

where x = rate of output (single ware production), v = vector of inputs, G(x) a monotonically increasing function, and g(v,t) homogeneous of degree 1 in v.

As regards the generation of the actual data several schemes can be envisaged. One hypothesis is that the production structure is of the putty-clay type (Johansen [1972]) with simple Leontief (limitational) ex post functions. To simulate the actual performance of plants an efficiency term with respect to the utilization of the inputs distributed in the interval (0,1) can be introduced multiplicatively on the r.h.s. of Eq. (1). We will adopt this scheme and in addition assume that the plants are operated on the "efficient corners" of the isoquants. Ex post the plant managers can only choose the rate of capacity utilization. With these assumptions concern about "slack" in fulfilling marginal conditions with respect to inputs is not relevant.

As regards the estimation procedure a key question is whether a specific distribution of the efficiency terms is assumed or not. If sufficient information is available to postulate a specific distribution the natural procedure is to derive maximum likelihood estimates as pointed out in Afriat [1972]. Without a specific efficiency distribution there are several ways to formulate the estimation problem as analyzed in Afriat [1972]. In this paper we will follow this latter approach. (Specific efficiency distributions will be pursued in a forthcoming paper.)

A natural objective -- with the information available -- is that the observations should be close to the frontier in some sense. In order to keep the estimation problem as simple as possible it is here chosen to minimize the simple sum of deviations from the frontier with respect to input utilization after logarithmic transformation, subject to on or below frontier constraints.

As regards the form of the production function the following specification is employed (below called the Zellner-Revankar, Z/R, specification, cf. Zellner-Revankar [1969]):

$$x^{\alpha} e^{\beta x} = A e^{\gamma_3 t} \cdot \prod_j v_j^{a_j + \gamma_j t} \tag{2}$$

Technical change is accounted for by specifying the possibility of changes in the constant term, A and the kernel elasticities, a_j for labour, L, and capital, K. With this specification the estimation problem turns out to be a standard linear programming problem. The objective function to be maximized becomes:

$$\sum_{t=0}^{T} \sum_{i=1}^{n} \{\beta \cdot x_i(t) + \alpha \cdot \ln x_i(t) - \ln A - \gamma_3 t$$
$$- (a_1 - \gamma_1 t) \ln L_i(t) - (a_2 + \gamma_2 t) \cdot \ln K_i(t)\} \tag{3}$$

The signs of the trends are preselected to the most probable signs. (This is unnecessary from a LP-technical point of view.) Note that although the objective function is linear in all the unknown parameters, the specification yields satisfactory flexibility as regards technical change.

The reader should observe that this is a deterministic calculation of the frontier. Its calculated parameters cannot be given a traditional stochastic interpretation.

Concerning the constraints of the LP-model, the expression within the brackets in (3) constitutes $(T+1)\cdot n$ constraints securing the observed input points to be on or below the frontier:

$$\beta \cdot x_i(t) + \alpha \cdot \ln x_i(t) - \ln A - \gamma_3 t - (a_1 - \gamma_1 t) \cdot \ln L_i(t)$$
$$- (a_2 + \gamma_2 t) \cdot \ln K_i(t) \leqq 0; \quad i=1,\dots,n; \; t=0,\dots,T. \tag{4}$$

In addition, we have the homogeneity constraint

$$\sum_j a_j(0) = 1. \tag{5}$$

Since there are only two trends in the kernel function Eq. (5) implies the restriction:

$$-\gamma_1 t + \gamma_2 t = 0; \qquad t = 1,\ldots, T \tag{6}$$

In addition, we want the kernel elasticities including trends to be restricted to the interval (0,1):

$$a_1 - \gamma_1 \cdot T \geqq 0 \tag{7}$$

$$a_2 + \gamma_2 \cdot T \leqq 1 \tag{8}$$

Finally we have the restrictions

$$\beta, \alpha, a_1, a_2, \gamma_1, \gamma_2, \gamma_3 \geqq 0 \tag{9}$$

(Note that ln A is not restricted to non negative values.)

3. THE ESTIMATION OF THE AVERAGE FUNCTION

With the assumptions adopted in this paper some care must be taken concerning the interpretation of an average function. It serves here only the function of giving an "average" picture of the ex post relationship between inputs and outputs across plants operating with different fixed input coefficients and capacity levels. The average function is specified to have the same functional form as the best-practice function shown in Eq. (2). (Note that the scale function is assumed to be unchanged over time.) This facilitates an analysis of structural change, but it must be noted that such an AP-specification must only be interpreted as convenient approximation to the actual relationship generated by adding new capacity according to the estimated BP-function.

As regards the estimation procedure we now start out with the assumption that deviations from the average function are simulated by introducing a random variable $N(0,\sigma)$, replacing the efficiency term in the BP-function. Maximum likelihood estimates are then obtained by using the adapted non-linear Box & Cox [1964] method outlined in Førsund [1974]. The essence of the method is to estimate the parameters on the r.h.s. of Eq. (2) after logarithmic transformation by OLS for trial values of α and β until a maximum of the likelihood function in question is reached.

4. THE DATA

In the empirical part of this study we have utilized primary data for 28 individual dairy plants during the period 1964-73. We have received all data from SMR (Svenska mejeriernas riksförening), a central service organization for the dairies in Sweden.

The milk production process can be divided into two stages: general milk processing, and packaging. The data refer to the first important stage in the milk production process, namely general milk processing. It includes the reception from cans or tanks of all milk, its storage and processing including pasteurizing and separation. Normally this stage defines the capacity of the plant. It is often treated as a separate unit by dairy engineers when discussing e.g. economies of scale and other aspects of costs.

Milk is regarded as a homogeneous product which is a very realistic assumption (in a very literal sense; milk is homogenized). Thus output is measured in tons of milk delivered to the plant each year. The amount of milk received is equal to the amount produced. There is no measurable waste of milk at this stage. According to SMR any difference is due to measurement errors. (Differences were of the magnitude of kilos.)

The labour input variable is defined as the hours worked by production workers including technical staff usually consisting of one engineer.

Capital data of buildings and machines are of the user-cost type, including depreciation based on current replacement cost, cost of maintenance and rate of interest. They have been centrally accounted for by SMR according to the same principles for all plants and after regularly capital inventory and revaluations of engineers from SMR. Note that this capital measure is proportional to the replacement value of capital, which can serve as a measure of the volume of capital, (see Johansen & Sørsveen [1967]). As regards the central question of capacity utilization we have investigated a measure based on the monthly maximum amount of milk received compared with the yearly average. This ratio is fairly stable for each plant over time, and the differences between plants are not very great. In consequence we have not corrected for capacity utilization. The increasing output over time for most of the plants supports the assumption.

5. THE EMPIRICAL RESULTS

The estimates of the parameters of the frontier and average production function are shown in Table 1 and the figures below. As the table reveals the trends in the marginal elasticities are important. In best-practice the trend in A is zero but becomes negative in average practice. Optimal scale obtains a considerably higher value in average practice than in best practice. The output of the largest plant has been in the interval 111 000 - 141 000 tons in the period 1964-73, except 1965 when it was 77 000, while the average output has increased from 29 000 to 39 000 tons. Taking our results at face value there are gains to be riped by increasing the average size of plants, but the gains are exaggerated by the average function.

Table 1. Estimates of the best-practice and the average practice production function. Combined time-series cross-section analysis. Estimates of the parameters of the production function

$$x^{\alpha}e^{\beta x} = Ae^{\gamma_3 t}L^{(a_1-\gamma_1 t)} \cdot K^{(a_2+\gamma_2 t)}$$

(t = 0 in 1964, t = 9 in 1973)

	Constant term ln A	Trend A $\gamma_3 \cdot 10^2$	Labour elasticity $a_1-\gamma_1 t$ 1964	1973	Trend L $\gamma_1 \cdot 10^2 = \gamma_2 \cdot 10^2$	Capital elasticity $a_2+\gamma_2 t$ 1964	1973	α	$\beta \cdot 10^5$	Optimal scale tonnes
Best practice	-8.17	0	.70	.41	3.14	.30	.59	.13	1.7	52 122
Average practice	-3.14	-6.83	.69	.37	3.62	.31	.63	.69	.40	76 610

The shape of the production functions and their development through time are shown in Fig. 1. Cutting the production functions with a vertical plane through the origin along a factor ray one obtains the classical textbook S-shaped graph of the frontier and average production function.

When assessing the somewhat surprising result above one should note the possibilities of systematic biases with the two estimation methods. Fig. 1 shows that the BP-function lies below the AP-function for small levels of output (no observations are, in fact, in this range). The BP-function is placed as close to the observations as possible, observing the on or above restrictions, including the observations of the smallest plants. The AP-function cuts through the observations of the middle range plants and lifts over the smallest while the BP-function has to be more curved in order to obey the restrictions when minimizing the sum of deviations. If engineering information could be obtained it might well turn out that it is a misspecification to allow the smallest plants to be on the frontier.

The characteristics of technical advance can also be illustrated in the input coefficient space (cf. Salter [1960] chapter 3) by the development of the technically optimal scale curve which we will call the efficiency frontier in the case of the best practice function. See Figure 2. The efficiency frontier is the locus of all points where the elasticity of scale equals one, (see Frisch [1965] chapter 8), i.e., it is a technical relationship between inputs per unit of output for production units of optimal scale. Thus the efficiency frontier represents the optimal scale of the frontier production function. (In Johansen [1972] p.21 the efficiency frontier is referred to as the technique relation.) In the input coefficient space the frontier or ex ante production function defines the feasible set of production possibilities while the technique relation defines the efficiency frontier towards the origin of this set. (This consideration has been elaborated in detail in Førsund [1971].)

In Figure 2 the labour saving bias of technical progress is reflected in the change of the optimal scale curve and the efficiency frontier. Changes of milk reception from cans to tanks and self-cleaning separators together with one storey buildings are elements of this process of technical advance, and examples of labour saving techniques.

In average practice the trend in A gets a negative sign. In spite of this Figures 1 and 2 show that the average production function shifts upwards and that the optimal scale curve moves rapidly towards the ordinate axis and the origin, even though the optimal scale function is constant.

Note also that in spite of a higher optimal scale in the average function the efficiency frontiers are strictly closer to the origin and the axis than the corresponding optimal scale curves.

A comparison between Figures 1 and 2 illustrates two different aspects of technical progress; on the one hand the development

Figure 1. The change in the frontier and average production function through time.

Combined time-series cross-section analysis. The production function cut with a vertical plane through the origin along a ray, $(\mu L^{o}, \mu K^{o})$, $L^{o} = 15\ 000$ and $K^{o} = 200\ 000$

$$x^{\alpha} e^{\beta x} = Ae^{\gamma_3 t} (\mu L^{o})^{a_1 - \gamma_1 t} \cdot (\mu K^{o})^{a_2 + \gamma_2 t}$$

The factor ratio corresponds to OA in Figure 2.

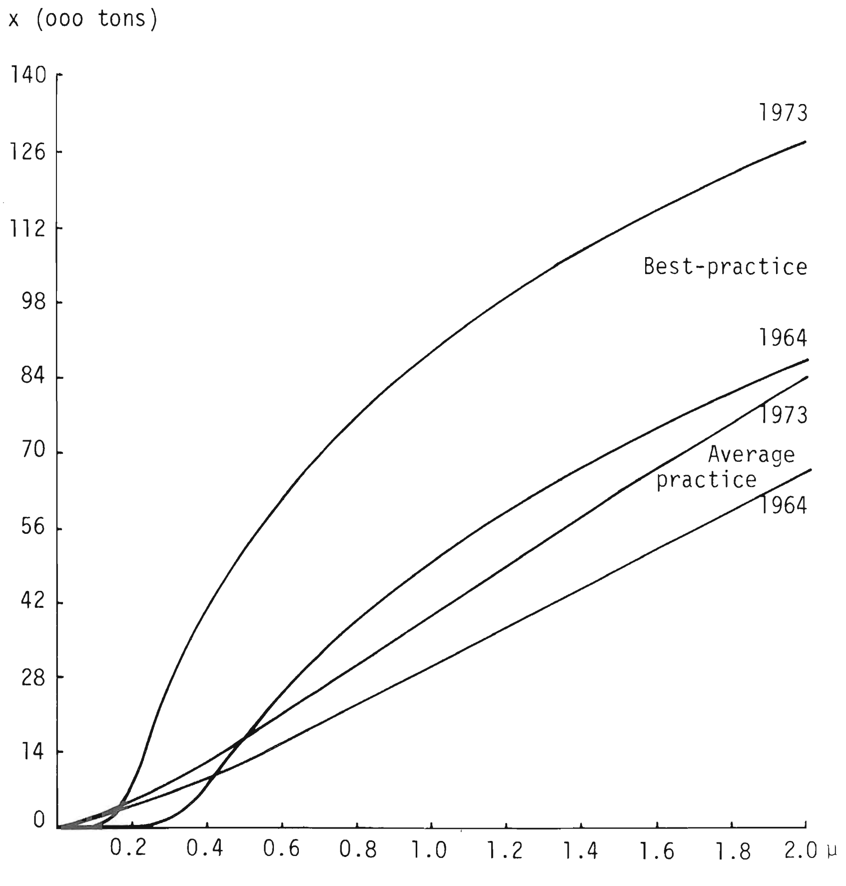


Figure 2. The changes in the efficiency frontier and the average optimal scale curve through time

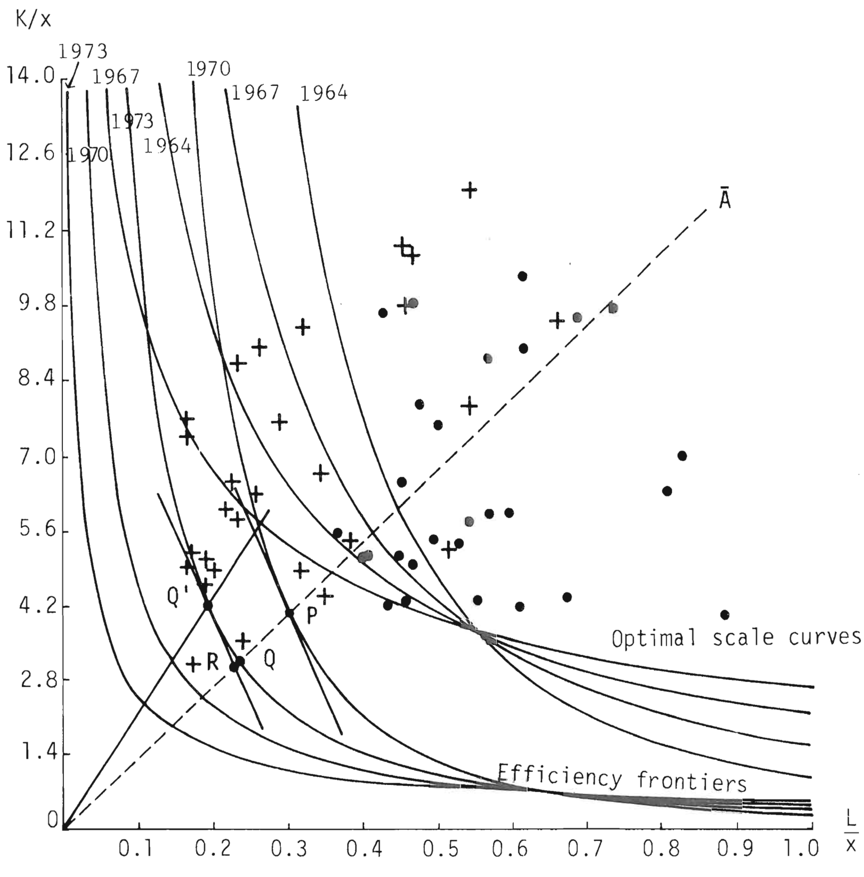
Combined time-series cross-section analysis. Estimates of the production function

$$x^{\alpha} e^{\beta x} = Ae^{\gamma_3 t} L^{(a_1-\gamma_1 t)} \cdot K^{(a_2+\gamma_2 t)}$$

with the efficiency frontier and optimal scale curve

$$\left(\frac{L}{x}\right)^{a_1-\gamma_1 t} \left(\frac{K}{x}\right)^{a_2+\gamma_2 t} \cdot Ae^{\gamma_3 t} \cdot \left(\frac{e\beta}{1-\alpha}\right)^{\alpha-1} = 1$$

The observed input coefficients for the years 1964 (dots) and 1973 (crosses) are plotted.



of the efficiency frontier and the optimal scale curve, on the other hand the development of the production function surface for a given factor ratio. While the most scale efficient plants are close to the efficiency frontiers, the best-practice production function reveals the most technically efficient plants which comprise both small and large plants, i.e., also scale inefficient plants. (These efficiency aspects will be treated in a separate paper. Se also Førsund and Hjalmarsson [1974].)

Measured along rays through the origin the distance between the efficiency frontier and the optimal scale curve has increased for all relevant factor ratios. Figure 1 also indicates that the distance between best-practice and average practice has increased during the period.

A numerical measure of the distance between best-practice and average practice can be obtained in several ways. (Førsund and Hjalmarsson [1974].) One measure utilized here is obtained by comparing the observed average output with the output obtained on the frontier function for the observed average amount of inputs. This measure can be regarded as a measure of structural efficiency and is denoted by S* and calculated according to the formula

$S^* = \frac{\bar{x}^o}{\bar{x}^*}$, where $\bar{x}^o$ is observed average production and $\bar{x}^*$ is obtained as the solution of $x^{\alpha}e^{\beta x} = A \cdot \prod_j \left(\frac{1}{n} \Sigma v^o_{ij}\right)^{a_j}$.

In the same way the distance between the average plant and the average function, $\bar{S}$, can be obtained.

A measure, S, which measures the distance between the frontier and average function can now be obtained by dividing S* with $\bar{S}$.

The numerical values of all three measures are presented in Table 2.

Table 2. The numerical values of S, S*, $\bar{S}$

Year	S	S*	$\bar{S}$
1964	0.60	0.61	1.01
1965	0.55	0.59	1.08
1966	0.53	0.54	1.01
1967	0.50	0.51	1.01
1968	0.54	0.51	0.95
1969	0.49	0.49	1.01
1970	0.47	0.46	0.97
1971	0.46	0.47	1.03
1972	0.42	0.47	1.12
1973	0.43	0.45	1.04

A clearly decreasing trend in the values of structural efficiency can be observed. One positive reason for this is a rapid technological progress which has increased the dispersion of the structure and the distance between the best practice and average practice techniques. All plants in the sample have survived the entire period. During the same time a lot of dairies have been closed down in Sweden. Thus the development of structural efficiency for all plants might have been another than for the set utilized here.

In order to improve the understanding of the technical change as measured in Figures 1 and 2 we will follow Salter's [1960] suggestions. He introduces three measures describing technical advance:

i) Rate of technical advance measured by the relative change in total unit cost for constant input prices;

ii) Labour- or capital saving bias measured by relative change in the optimal (cost minimizing) factor proportion for constant input prices;

iii) Relative change in the elasticity of substitution.

It might be of interest to note that the two first measures have direct connections with the overall- technical-, and price efficiency measures introduced by Farrell [1957].

When working with non-homogeneous production functions it is natural to replace the unit isoquants in Farrell's and Salter's analysis with the efficiency frontiers or scale curves shown in Figure 2. (See Førsund [1974], Førsund and Hjalmarsson [1974] for interpretations of the Farrell measures in a setting of inhomogeneous functions.) Let P in Figure 2 be the point of reference on the efficiency frontier for the base period. Q' is the point on the efficiency frontier for a later period where the marginal rate of substitution is the same. A measure analogous to the Salter measure i) above, assuming cost minimization, is then 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 realizing optimal scale. (In our case the optimal scale output is constant for the BP- and AP-functions.) This change is equal to OR/OP in Figure 2 which is also the Farrell overall efficiency measure for a production unit with observed input coefficients given by P relative to the next periods efficiency frontier.

The Farrell overall measure, and correspondingly the Salter technical advance measure, can be split multiplicatively into technical efficiency, OQ/OP, and price efficiency, OR/OQ. In our context this splitting shows the relative reduction in unit cost due to the movement along a factor ray and the movement along the next period efficiency frontier generated by biased technical change.

The efficiency frontiers or scale curves used here are given by

$$\xi_{2,t} = A^{-1/a_{2,t}} e^{-\gamma_3 t/a_{2,t}} \left(\frac{e\cdot\beta}{1-\alpha}\right)^{(1-\alpha)/a_{2,t}} \cdot \xi_{1,t}^{-a_{1,t}/a_{2,t}} \tag{10}$$

where $\xi_1 = L/x$ and $\xi_2 = K/x$. The marginal rate of substitution MRS, for this function is equal to the MRS for the production function and equal to

$$\frac{-d\xi_{2,t}}{d\xi_{1,t}} = \frac{a_{1,t}}{a_{2,t}} \cdot \frac{\xi_{2,t}}{\xi_{1,t}} \tag{11}$$

Salter's measure of bias is, in general:

$$D_{21} = \frac{\xi_{2,t+1}/\xi_{1,t+1}}{\xi_{2,t}/\xi_{1,t}} \tag{12}$$

when keeping factor prices constant, or equivalently, keeping the MRS constant. We then get:

$$\frac{a_{1,t}}{a_{2,t}} \frac{\xi_{2,t}}{\xi_{1,t}} = \frac{a_{1,t+1}}{a_{2,t+1}} \frac{\xi_{2,t+1}}{\xi_{1,t+1}}, \text{ i.e.}$$

$$\frac{\xi_{2,t+1}/\xi_{1,t+1}}{\xi_{2,t}/\xi_{1,t}} = \frac{a_{1,t}/a_{2,t}}{a_{1,t+1}/a_{2,t+1}} . \tag{13}$$

Since the elasticity of substitution is constant and equal to 1 the relative change in the factor ratio (the MRS being constant) is equal to the relative change in the MRS for a constant factor ratio, $b = \xi_2/\xi_1$:

$$\frac{MRS_t}{MRS_{t+1}} = \frac{a_{1,t}}{a_{2,t}} \cdot b \bigg/ \frac{a_{1,t+1}}{a_{2,t+1}} \cdot b = \frac{a_{1,t}/a_{2,t}}{a_{1,t+1}/a_{2,t+1}} . \tag{14}$$

Note that the bias measure is here independent of the price- or factor ratio chosen.

The Salter technical advance measure, choosing the Laspeyre version for convenience, becomes:

$$T = (\xi_{1,t+1} + \xi_{2,t+1}\left(\frac{\partial x/\partial K_t}{\partial x/\partial L_t}\right) \Big/ \xi_{1,t}+\xi_{2,t}\left(\frac{\partial x/\partial K_t}{\partial x/\partial L_t}\right) = \frac{\xi_{1,t+1}}{\xi_{1,t}} \cdot \frac{a_{1,t}}{a_{1,t+1}} \qquad (15)$$

utilizing that $MRS_t = MRS_{t+1}$ and that the kernel function is homogeneous of degree 1. We find it more convenient here to start out from a given factor ratio, $b = \xi_2/\xi_1$, rather than a price ratio. (This is, of course, equivalent.) From (10) we then have

$$\xi_{1,t} = b^{-a_{2,t}} A^{-1} e^{-\gamma_3 t} \left(\frac{e\beta}{1-\alpha}\right)^{1-\alpha} \qquad (16)$$

where b is the chosen factor ratio. Remembering (13) yields

$$\xi_{1,t+1} = (D_{21} \cdot b)^{-a_{2,t+1}} A^{-1} e^{-\gamma_3(t+1)} \left(\frac{e\beta}{1-\alpha}\right)^{1-\alpha} \qquad (17)$$

Inserting (16) and (17) in (15) introducing $a_{1,t} = a_1 - \gamma_1 t$, $a_{2,t} = a_2 + \gamma_2 t$ yields

$$T = e^{-\gamma_3} b^{-\gamma_2} D_{21}^{-a_2-\gamma_2(t+1)} \cdot \frac{a_1-\gamma_1 t}{a_1-\gamma_1(t+1)} . \qquad (18)$$

The relative unit cost reduction due to a movement along a factor ray (Farrell technical efficiency) is

$$(\xi_{1,t+1}/\xi_{1,t})_{b=\text{const.}} = e^{-\gamma_3} b^{-\gamma_2} \qquad (19)$$

The price- or allocative efficiency measure must then be

$$D_{21}^{-a_2-\gamma_2(t+1)} \cdot \frac{a_1-\gamma_1 t}{a_1-\gamma_1(t+1)} . \qquad (20)$$

We see the close connection between the relative unit cost reduction due to the bias and the Salter bias measure with our functional specification. The "pure movement" measure, OQ/OP, is here independent of time, but depends on the chosen factor ratio (relative factor prices) and the trend parameters, while the bias gain measure, OR/OQ, is independent of the factor ratio (relative factor prices), but depends on time and the bias trend parameter.

The various measures corresponding to the estimates reported in Table 1 are set out below.

Table 3. Characterization of technical change by the movements of the efficiency frontiers and optimal scale curves,[a] Salter measures and Farrell-inspired splitting-up
Factor ratio b = 13.33 corresponding to OA in Figure 2

Type of measure	AP		BP	
	1964-65	1972-73	1964-65	1972-73
Technical advance:				
Overall relative change in unit cost on optimal scale	.9719	.9722	.9198	.9200
Proportionate unit cost change	.9749	.9749	.9219	.9219
Bias unit cost change	.9970	.9972	.9977	.9980
Labour saving bias:				
Relative change in capital-labour ratio	1.1786	1.1658	1.1565	1.1365

[a] Note that since we operate with constant scale functions the measures in Table 3 are independent of the output level chosen.

The splitting-up of the total reduction in unit cost reveals that although the yearly optimal increase of the capital-labour ratio is about 17 % for the AP- and 15 % for the BP-function this change yields minimal cost reductions, .3 to .2 %. It is the displacement of the frontier towards the origin as measured along a factor ray (Farrell technical efficiency of the past technology relative to the present) that results in significant reductions in unit costs; about 3 % for the AP- and 8 % for the BP-function. The AP-function has a somewhat stronger labour saving bias and a markedly slower displacement of the optimal scale curve towards the origin than the BP-function.

One possible economic explanation of this sustained difference is that the total capacity of the sector has been increasing, at a yearly average of 3.34 % only, implying an investment growth rate too small to update average sector performance in pace with best-practice performance.

Another explanation might be that technical progress is overestimated by the frontier function during the last years of the period because we have assumed constant trends during the whole period. (The development of the marginal elasticities must be broken sooner or later as the values are restricted to the interval (0,1). During the whole period five plants is on the frontier, two year 0, one year 1, one year 4 and one year 8. Thus in the last year no plant is on the frontier and the slacks show that the distance to the nearest plant is relatively large. On the other hand the next last year one plant is on the frontier.

REFERENCES

Afriat. S.N. ,1972, "Efficiency Estimation of Production Functions". International Economic Review, 13(3), October, pp.568-598.

Aigner, O.J. and Chu, S.F., 1968, "On Estimating the Industry Production Function". American Economic Review, 58(4), pp. 226-239.

Box, G.E.P. and Cox, D.R., 1964, "An Analysis of Transformation". Journal of the Royal Statistical Society, (Series B), 26(2), 211-245.

Carlsson, B., 1972, "The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries 1968". Swedish Journal of Economics, 74(4), pp.468-485.

Farrell, M.J., 1957, "The Measurement of Productive Efficiency". Journal of the Royal Statistical Society, Series A (General), Part III, 120, pp.253-290.

Frisch, R., 1965, Theory of Production. D. Reidl Publishing Co. Dordrecht-Holland.

Førsund, F.R., 1971, "A note on the technically optimal scale in inhomogeneous production functions". Swedish Journal of Economics, 73(2), pp.225-240.

Førsund, F.R., 1974, "Studies in neoclassical theory of production". Memorandum from the Institute of Economics, University of Oslo, February 4.

Førsund, F.R., 1975, "The Homothetic Production Function". Swedish Journal of Economics, 77(2), pp.234-244.

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

Johansen, L. and Sørsveen, Å., 1967, "Notes on the measurement of real capital in relation to economic planning models". The Review of Income and Wealth, 13(2), June, pp.175-197.

Johansen, L., 1972, Production Functions, North,Holland, Amsterdam.

Meeusen, W. and van den Broeck, J., 1976, Efficiency estimation from Cobb-Douglas production functions with composed error. Paper presented at the European Meeting of the Econometric Society, Helsinki.

Salter, W.E.G., 1960, Productivity and Technical Change. University Press, Cambridge.

Timmer, C.P., 1971, "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency". Journal of Political Economy, 79(4), pp.776-794.

Zellner, A. and Revankar, N.S., 1969, "Generalized Production Functions". Review of Economic Studies, 36(2), April pp.241-250.

PRODUCTION FRONTIERS OF INDIVIDUAL FIRMS IN SWEDISH MANUFACTURING 1975 AND 1976

Jim Albrecht
Columbia University, New York

I. INTRODUCTION

This paper describes how data from Industrial Planning Surveys conducted by the Federation of Swedish Industries will be integrated into the "micro-to-macro simulation model of the Swedish economy" (MOSES) that is being developed at IUI under the direction of Gunnar Eliasson. It is intended to be a preliminary reference paper for the production system ascribed to the MOSES production units and to catalogue the related data that are generated within their real-world Swedish counterparts. As such it may be of use both to those who are interested in how the MOSES production block functions and to those who are interested in production planning within large industrial firms.

These surveys, which have been designed since 1975 with MOSES in mind, are sent out annually to the 250 largest Swedish industrial firms. Since these surveys focus on production planning, rather than financial planning, they are conducted at the production-planning level. For many of the relatively small firms the firm level and the production-planning level are identical; however, in the larger, multi-product firms these are often not the same, so questionnaires are sent out to as many as 7 or 8 divisions within some firms. Data referring to strictly domestic Swedish operations are available in the following areas:

(1) employment and wage bill
(2) sales (abroad and domestic)
(3) raw materials purchases
(4) investment expenditures (both plant and equipment)
(5) production volume

(6) capacity utilization
(7) orders
(8) inventories (product and raw materials)

The data on employment and the wage bill are available for the year of the survey and, retrospectively, for the previous year. The data on sales, on raw materials purchases and on investment expenditures are available for the year of the survey, for one year retrospectively and for one year prospectively. The production volume data are given as per cent changes from the year prior to the survey to the survey year and from the survey year to the next year (as an expectation). Data on capacity utilization are quite unique and, since they are of central importance to the MOSES production block specification, are described in detail below. The orders data give a per cent change as compared with the year prior to the survey and give information about orders as a fraction of planned production. The inventories data give the "normal" and survey year ratios of product and raw material stocks (as of the end of the survey year) to sales volume and purchases of raw materials, respectively. Finally, supplementary questions are asked in each survey year. In 1975 and 1976 respondents gave information about impediments to investment, and in 1977 information was obtained about depreciation rates and about the relationship between investment and incremental capacity.[1]

In the sequel I first sketch the place of the model production system in the model as a whole. Then the specification of the production system is given in two parts -- the specification of a short run relationship between output and employment and the specification of the process whereby this short run frontier shifts between quarters. Finally, the algorithm for numerically fitting short run frontiers for individual firms is presented in detail. An algorithm for numerically specifying the shift mechanism and

---

[1] This description is only strictly correct for the 1976 and 1977 surveys. Certain details -- most crucially, information on changes in raw material inventories -- are omitted from the 1975 survey. A complete description of the data through the 1976 survey can be found in a paper by Ola Virin [1976].

its relationship to past investments is not yet available. This is because a time series of data of sufficient length has just become available and we do not yet know how robust various proposed algorithms are to imperfections in the data. I do, however, present graphical examples of actual shifts in production frontiers in an appendix to illustrate the basic considerations and numerical magnitudes involved.

II. SOME BACKGROUND ON MOSES

The MOSES production system described below needs to be understood within the context of the full model. It is obviously impossible to give a self-contained description of the full model in a short paper. However, a brief sketch should suffice to place the model production system in perspective. To obtain more information about the model one may consult a series of papers by Eliasson -- Eliasson with Heiman and Olavi [1976] gives the documentation for the simplest version of the model and Eliasson [1977, 1978] present applications of the model. In addition there is an IUI-IBM conference volume [1978] dealing with MOSES.

MOSES (for MOdel of the Swedish Economy in Simulation) is a "micro-to-macro simulation model of the Swedish economy". It is a <u>micro model</u> in the sense that the behaviors of certain individual economic entities (viz., industrial production units) are modelled in detail. It is a <u>micro-to-macro model</u> in the sense that these micro agents are embedded in the framework of a macro economy. It is a <u>micro-to-macro simulation model</u> in the sense that the equations of the model are solved recursively by computer. (The model language is APL.) And it is a <u>micro-to-macro simulation model of the Swedish economy</u> in the sense that the Swedish economy is taken as the object of description and the benchmark for calibration.

Unlike most micro-simulation models of macroeconomies, the primary purpose of MOSES is not that it be useful as a forecasting tool. Rather, the main idea is that the model be useful as a theoretical tool for understanding how the macroeconomy operates. In particular, the model was constructed with the idea of studying "the micro basis for inflation" and "the interactions between inflation, profitability and growth". The focus of the model is on the supply side of the macroeconomy, and this is probably in contrast to most analytical theorizing about the same subject.

The operation of the MOSES economy is represented by the operation of a sequence of modules, and the completion of one sequence represents one calendar quarter. The organization of these modules in the simplest MOSES economy can be seen in Figure 1.

At start-up the industrial portion of the MOSES economy is populated by a number of firms (from 30 up to, say, 1000) divided into four sectors. Each of these firms has a past history, and on the basis of this history forms single-valued expectations about sales, prices and wages and sets a profit-margin target (in modules EXP and TARG in Figure 1). Each firm also faces a short-run production possibilities constraint giving the maximum output attainable for any level of employment. Both output and labor are homogeneous. The PRODPLAN module combines this production possibilities constraint with the firm's expectations and profit-margin target to produce a preliminary output/employment plan. The process by which this plan is set is based on interview studies conducted by Eliasson [1976] and may be characterized as one of satisficing. The basic algorithm can be illustrated using Figure 2.

Figure 2 charts output (Q) against labor (L). The short run production possibilities frontier is given by the function QFR(L), and a (Q, L) combination such that $Q \leq QFR(L)$ is said to be feasible. The particular feasible output/employment plan that is chosen depends upon the satisficing criterion. A (Q,L) combination that satisfies the firm's profit-margin target conditional on the price

Figure 1. Modular Structure of the Simplest MOSES-Economy

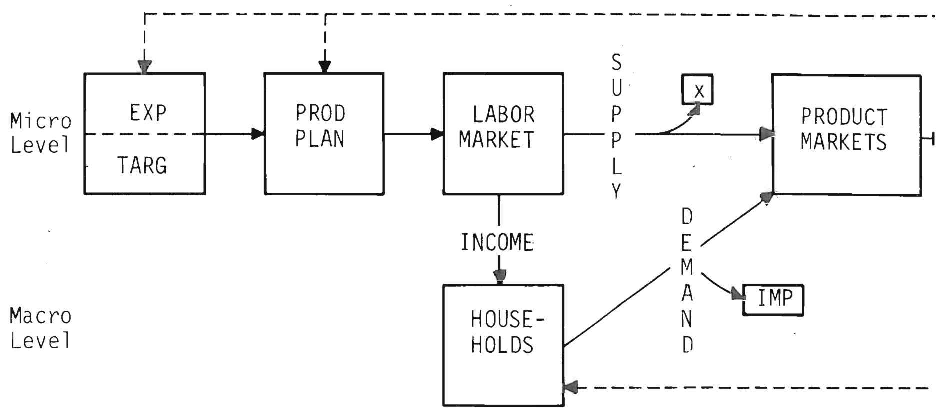
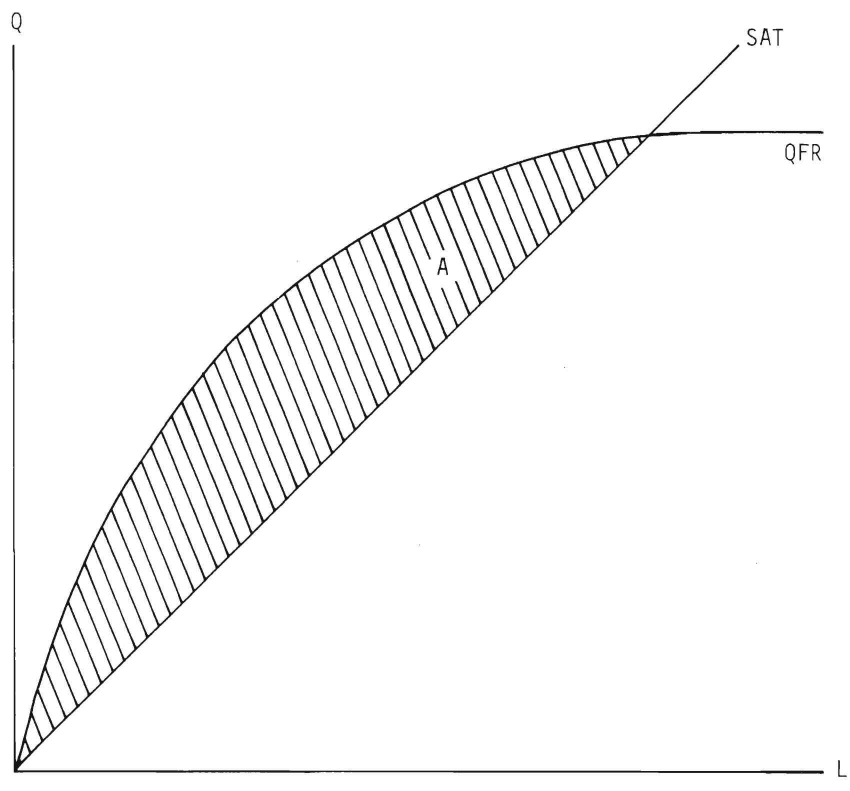


Figure 2. PRODPLAN Module



it expects to receive for its output and on the wage rate it expects to pay its labor is said to be satisfactory, and the ray SAT divides the (Q,L) plan into satisfactory versus unsatisfactory output/employment plans. The set of simultaneously feasible and satisfactory (Q,L) plans is thus illustrated by the lens area in Figure 2. Given this framework the PRODPLAN choice algorithm can be described as a rule to specify a trial set of (Q,L) plans (based on the firm's retained labor force and on its expected sales deflated by the expected price adjusted for a range of desired inventory change) and a group of rules to adjust this trial set if it does not intersect the lens area. The result of this algorithm will be a trial (Q,L) plan, e.g., point A in Figure 2, which may or may not be on the short run production possibilities frontier.

Upon completion of the PRODPLAN module, each firm has a planned labor force and a planned level of production, but these plans may not be feasible in the aggregate. Firms must confront one another and interact with the consuming public to resolve any inconsistencies, and the remainder of the quarterly module sequence may be thought of as a process of harmonizing firms' production plans.

The first confrontation takes place in the LABOR MARKET. Should any firms' plans call for the hiring of additional labor, some recruitment must be carried out either from a pool of unemployed workers or by raiding other firms. This process of recruitment and raiding produces an employment level and a wage rate for each firm, implying an aggregate wage bill which then goes to households (specified in the macro) as income. This income then becomes an argument of the aggregate demand system ascribed to households (based on macro estimates of a modified linear expenditure system with habit formation from Dahlman and Klevmarken [1971]). An additional output of the LABOR MARKET module is a supply and price quotation from each firm. A firm's supply will differ from its output plan if it has been unable to meet its recruitment plan. These firm supplies and price quotations are then aggregated to produce an aggregate supply and an initial price offering on each of the four markets corresponding to the four industrial sectors.

The final confrontation then takes place in the PRODUCT MARKETS module. After adjusting aggregate supply for exports (X) and aggregate demand for imports (IMP), aggregate supply and aggregate demand (as a function of households' income, the vector of price offerings, etc.) are compared on each of the four markets, and prices are adjusted according to the sign and magnitude of excess demand. This process continues through a pre-determined number of iterations, resulting in final prices and final sales on each of the four markets. Final sales are then spread across the production units comprising each of the four industrial sectors, and inventory change is computed as the residual between production and sales on the individual firm level.

Thus, after the completion of the quarterly sequence of modules pictured in Figure 1, each model firm has realized a level of sales, a price for its product, a level of inventory change, a wage rate, a level of employment and, by simple computation, a profit margin. The realized sales, price, wage rate and profit margin are fed back into the EXP and TARG modules as the newest component of past history, and the new levels of inventories and employment are fed back into the PRODPLAN module. Likewise, the newly generated consumption history is fed back into the HOUSEHOLDS module to become an argument of aggregate demand.

This completes the link between quarters in the simplest version of MOSES with one important exception, viz., the updating of the production possibilities frontier via investment. In this simplest version investment is equated (approximately) to ploughed-back profits, and the PRODPLAN module has an algorithm that relates the shift in QFR(L) to new investment.

Finally, it should be emphasized that the model description given above is grossly simplified; in particular, no mention has been made of the service sector, of government taxation and expenditures, of intermediate goods, inter industry markets and input-output constraints, of long-term borrowing decisions and the financing of investment, and of the monetary sector. All of these are important for understanding the operation of the complete model but are of secondary importance for understanding the place of the firms' production systems in the model.

The short run analysis given below discusses an assumed parametric form for QFR(L). Given this assumed parametric form, a method for computing (estimating) QFR(L) for each firm in each year using the Industrial Planning Survey data is described. This method is based on the ex post observation of the point A and on the capacity utilization information supplied by respondents. In addition to their use in estimating the function QFR(L), these data allow one to check the capacity utilization figures produced by the model against their real-world counterparts. The "between quarters" analysis explains the algorithm relating shifts in QFR(L) to investment in some detail. However, a technique for numerically specifying this algorithm is not yet settled upon.

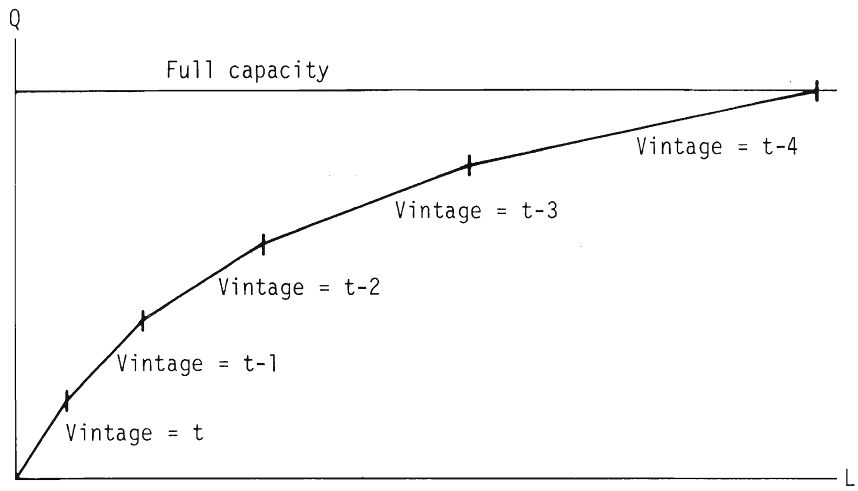
III. SPECIFICATION OF THE MOSES PRODUCTION SYSTEM

A. The Short Run Production Frontiers

The parametric form assumed for the production frontiers (QFR(L)) ascribed to the MOSES firms in the short run can be motivated by the following argument. Since we are specifying production possibilities on a very micro level, we want to express the production system in such a way that information collected at that micro level can be used in its specification. This means that it is necessary to eschew any use of the concept of a capital stock. The alternative is to use a vintage, putty-clay approach in which incremental capacity is produced by investment. Technological change is naturally introduced within such a framework by specifying the rate at which labor requirements per unit of incremental capacity decline over time. The function QFR(L) can thus be visualized as a chaining together of ex post (i.e., clay) relationships between incremental capacity and labor requirements with the ratio between the two determined by the vintage of the production process (cf. Figure 3). The convexity of the production possibilities set (i.e., the sequencing of the vintages) simply reflects the idea that as a firm contracts its operations towards the origin, the less efficient production processes will be the first to be shut down.

The problem with an explicit vintage approach is that it requires the storage of large amounts of information. We want to retain the vintage idea without retaining the informational requirements. The obvious remedy is to approximate the discrete formulation of QFR(L) by a continuous function expressed in terms of as few parameters as possible.

Figure 3. A discrete, vintage formulation of QFR(L,t)



Looking back to Figure 3, notice that the slope of any line segment is simply the ratio of incremental capacity to incremental labor requirements (i.e., $\Delta QFR/\Delta L$). The most straightforward way to introduce technical change is to make a simple assumption about how these ratios change (i.e., about how $\Delta QFR/\Delta L$ changes), and the assumption made is $\Delta^2 QFR/\Delta L^2 = -\gamma(\Delta QFR/\Delta L)$, i.e., a constant percentage increase in labor productivity. This constant rate, γ, is entered as a negative since $\Delta L > 0$ implies a movement to an older vintage.

The continuous formulation of this technological assumption is $d^2QFR(L)/dL^2 = -\gamma[dQFR(L)/dL]$, implying $[dQFR(L)/dL] = Ke^{-\gamma L}$, with K an arbitrary constant. This simply says that the marginal product of labor equals a constant multiple of $e^{-\gamma L}$. Simple integration then gives $QFR(L) = c - (K/\gamma)e^{-\gamma L}$, with c another arbitrary constant, or $QFR = c[1 - (K/c\gamma)e^{-\gamma L}]$. To fix the constants we first require that $QFR(0) = 0$, implying $K/c\gamma = 1$. Then we impose an upper bound on QFR(L) which for reasons explained below is written as QTOP·(1-RES); thus $c = \lim_{L\to\infty} QFR(L) =$ QTOP·(1-RES).

The short run production frontier is therefore $QFR(L) = QTOP\cdot(1-RES)(1-e^{-\gamma L})$; i.e., the continuous approximation to the discrete vintage formulation outlined above is expressed in terms of 3 parameters -- QTOP, RES and γ.

QTOP represents full capacity, i.e., the "maximum possible" output given the application of an unlimited amount of labor, and RES represents the fraction of this full capacity that is held as "reserve slack". The existence of such slack has been discussed in the organizational literature, and evidence on its existence has been presented in Eliasson [1976]. Firm management knows that this waste exists but does not know its magnitude, nor does it know how (operationally) to reduce it. This is not necessarily bad, however, since reserve slack can accumulate over time (up to a limit) and be available as a buffer should normal production planning procedures fail to yield a satisfactory plan.

Numerically, this presents a problem. Since reserve slack is "hidden" by definition, one cannot hope to directly elicit any information about its magnitude from businessmen. Likewise, one cannot hope to directly identify QTOP. Instead, one can only observe the product QTOP(1-RES), i.e. "normal capacity", in the data, but this is sufficient for a numerical specification of the "normal" short run production frontier.

It should be noted that once normal capacity is fixed, the parameter γ has a very straightforward interpretation. For a given QTOP(1-RES), γ parameterizes a family of curves approaching that asymptote. If two production possibilities frontiers with equal normal capacity are compared, the one with the smaller value of γ dominates in the sense that for any given level of employment the maximum attainable output is greater. That is, γ characterizes the efficiency with which a given capacity is approached.

B. Shifting the Production Possibilities Frontier

The function QFR(L) shifts "between quarters" because of depreciation and in response to the firm's past investments (INV). The depreciation assumption currently used in the model is the simplest one possible, namely that QTOP depreciates at a constant rate ρ per quarter. Questions about the "economic life expectancy" of plant and equipment were included in the 1977 questionnaires, and it is possible that the responses to these questions will allow direct estimation of ρ. This specification is, however, somewhat lacking in the vintage spirit, and some alternatives based on suggestions made in Bentzel's paper in this volume will be tried.

The effects of past investment are naturally more complicated. Investment affects all 3 parameters of QFR(L) -- QTOP, RES, and γ. The increase in QTOP (after allowance for depreciation) is approximately related to investment deflated by a durable goods price index by a fixed coefficient called INVEFF, i.e., ΔQTOP $\simeq$ INVEFF$\cdot$INV. The modifier "approximately" is needed because some of this incremental capacity is slated to go into reserve slack. If "enough" reserve slack already exists, then some of the effects of investment will be wasted. This process depends upon 2 parameters in the model: (1) LOSS -- the fraction of incremental capacity which is immediately diverted into reserve slack, and (2) RESMAX -- the maximum ratio of slack to full capacity, i.e., an upper bound on RES. Like RES itself, these 2 parameters are by definition unobservable, and survey questionnairing cannot shed any light on their magnitudes.[1]

In addition to its effects on capacity, investment can also increase the efficiency with which any given capacity is approached; i.e., investment will affect γ as well as QTOP. Define TEC = QTOP$\cdot\gamma$, and notice that $dQFR(L)/dL = (1-RES)\cdot TEC\cdot e^{-\gamma L}$, so that $dQFR(0)/dL = (1-RES)\cdot TEC$. Thus, TEC measures the labor productivity ("marginal product") of the production process of most recent vintage. It is natural to view investment as directly affecting TEC, and this is how the "productivity-enhancing" aspects of investment are treated in the model. TEC is updated by an exogenous factor MTEC in proportion to QTOP. Specifically, a harmonic averaging process has been used:

$$TEC(t) = QTOP(t) \Big/ \left(\frac{QTOP(t-1)}{TEC(t-1)} + \frac{\Delta QTOP}{MTEC} \right),$$

implying

$$\gamma(t) = 1 \Big/ \left(\frac{1}{\gamma(t-1)} + \frac{\Delta QTOP}{MTEC} \right).$$

The crucial parameter in this updating is clearly MTEC, and this is specified in the model as $MTEC(t) = MTEC(0)\cdot(1+QDMTEC)^t$. Thus,

---

[1] A full description of this process, including the details of how RES is updated, can be found on pp. 206-07 in Eliasson with Heiman and Olavi [1976].

QDMTEC is interpreted as the (quarterly) rate of technical change or the rate of change in the productivity of new investment.

The empirical problem in this section is to relate the pair (ΔQTOP,ΔΥ) to past investments. This should probably be approached as an econometric problem rather than one of simple numerical specification; i.e., it is probably more important to get a good single estimate of INVEFF to apply to all firms than it is to specify that parameter on an individual, firm by firm basis. The chief difficulty is to relate (ΔQTOP,ΔΥ) to the correct distributed lag of past investments, i.e., to specify the rate at which new investments become operational. In principle, this problem could be approached directly by asking respondents to specify lead times for typical investment projects, but so far this has not been done in these surveys.[1] This seems especially called for because the surveys only provide annual information on investment and on changes in the production frontier parameters, whereas MOSES is specified on a quarterly basis.

Since questions of time lags are involved an empirical approach has been postponed until a time series of data of sufficient length becomes available. The 1977 survey responses (just available as this is written) will be the first used to get at these problems. In the meantime INVEFF is specified for each firm as the ratio between its value added and the replacement value of its production capital, and the investment completion lag is specified as a 3rd-order exponential delay function with average delay time equal to TMINV (specified exogenously). The relationship between ΔΥ and investment is implicitly specified through QDMTEC. It should be noted that this specification of INVEFF is in accord with the simple assumption of a constant capital/output ratio (although here we are working with the current replacement value of production capital and <u>not</u> with the "volume" of the ca-

---

[1] There is some related evidence available in Mayer [1960]. It is surprising that so little effort has been expended towards directly estimating completion lags since these are "adjustment lags" of the sort that figure prominently in the popular "neoclassical models of investment".

pital stock) that is commonly adopted both in economic modelling and in firms' planning routines. Further, TMINV and QDMTEC are not so "arbitrarily" specified as it might first appear since these can be calibrated on the basis of total model performance.[1] However, the object of these "outside" empirical specifications is precisely to get the number of parameters which must be so calibrated down to an irreducible minimum.

IV. A METHOD FOR NUMERICALLY SPECIFYING QFR(L)

The numerical procedure for specifying QFR(L) is quite simple and is illustrated in Figure 4. Suppose a firm is observed producing at the point A = (L*,Q*). The first step in computing QFR(L) is to compute QTOP(1-RES) ≡ QLIM, so what is needed is a measure of total excess capacity, i.e., QLIM-Q* or the distance AC. Once QLIM is computed, the distance AB needs to be measured to generate the point QFR(L*). Then $QFR(L^*) = QLIM(1-e^{-\gamma L^*})$ is used to compute γ. What is used to make these computations is, in effect, a division of total excess capacity (QLIM-Q*) into a measure of "labor redundancy" (QFR(L*) - Q*) and a measure of "capital redundancy" (QLIM-QFR(L*)). But this division is precisely what the capacity utilization questions in the Industrial Planning Surveys are designed to elicit.[2]

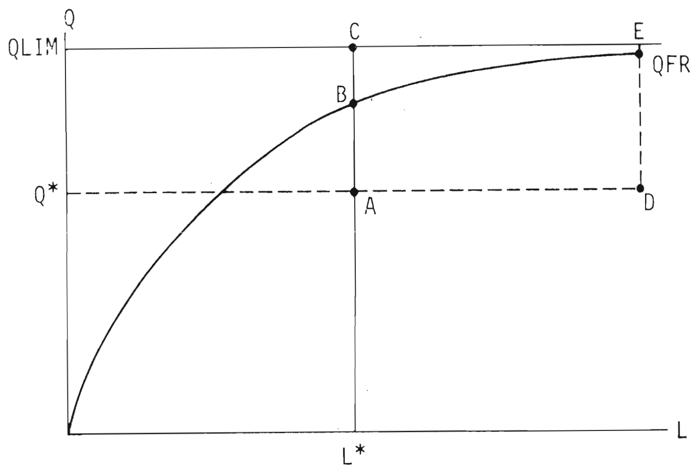
The first of these capacity utilization questions asks (loosely translating), "By what percentage could output be increased, product demand permitting, but with existing employment?" The second question asks "By what percentage could output be increased, product demand permitting, and employment as large as needed?" In terms of Figure 4, what this second question suggests is that firms think of expanding employment to some "very large" level,

---

[1] It is also possible to specify QDMTEC via interviews with production engineers. See Bo Carlsson's article in the IUI-IBM Conference volume (Eliasson (ed.) [1978]).

[2] Data on the division of unused capacity between labor redundancy and capital redundancy are presented in the Appendix.

Figure 4. Computing QLIM and γ



say point D. The question then is: What is the discrepancy between D and E?" These two percentages are called A21 and SUM in MOSES notation; thus,

QFR(L*) = Q*(1+A21) and

QLIM = Q*(1+SUM).

Then given L*, Q*, A21 and SUM, QFR(L*) can be inverted to solve for

$\gamma = \{-\ln[(QLIM-QFR(L^*))\div QLIM]\}/L^*$.

Notice that if A21 = SUM (i.e., no "capital redundancy"), this last calculation cannot be made.

The firms' employment levels (L*) are obtained directly from the surveys; however, there are some practical problems involved in specifying Q*. The survey gives information on sales, rather than value added, so some adjustments need to be made. In particular, we need to subtract off raw materials purchases and to allow for changes in inventories of both products and raw materials. Information on raw materials purchases is available for all 3 years (i.e., 1975-77), albeit retrospectively for 1975 (notation:

R75, R76 and R77). In the 1976 and 1977 surveys firms were asked to give the ratio of product inventories (STO) to sales (S) for the survey year (i.e., STO76/S76 and STO77/S77) and "normally" ($\overline{STO}/\bar{S}$); likewise firms provided the ratio of raw materials inventories (RSTO) to raw materials purchases for the survey year (RSTO76/R76 and RSTO77/R77) and "normally" ($\overline{RSTO}/\bar{R}$). In the 1975 survey firms were asked to specify the relative deviation of product inventories from their "normal" levels ($(STO75-\overline{STO})/\overline{STO}$), but no questions were asked about raw materials inventories.

As approximations, the following calculations are made:

$S75[(STO75 - \overline{STO})/\overline{STO}]\overline{STO}/\bar{S} \approx STO75-\overline{STO}$ is taken as a measure of the change in product inventories for 1975;
$S76[(STO76/S76) - (\overline{STO}/\bar{S})] \approx STO76 - \overline{STO}$ is taken as a measure of the change in product inventories for 1976 (and similarly for 1977);

$R76[(RSTO76/R76) - (\overline{RSTO}/\bar{R})] \approx RSTO76 - \overline{RSTO}$ is taken as a measure of the change in raw materials inventories for 1976 (and similarly for 1977).

Raw materials inventories are assumed unchanged for 1975. Value added for 1975 through 1977 (Q75, Q76 and Q77) is thus computed as:

$$Q75 = S75\{1 + [(STO75-\overline{STO})/\overline{STO}]\overline{STO}/\bar{S}\} - R75$$

$$Q76 = S76\{1 + [(STO76/S76) - (\overline{STO}/\bar{S})]\}-R76\{1-[(RSTO76/R76)-(\overline{RSTO}/\bar{R})]\}$$

$$Q77 = S77\{1 + [(STO77/S77) - (\overline{STO}/\bar{S})]\}-R77\{1-[(RSTO77/R77)-(\overline{RSTO}/\bar{R})]\}.$$

The problem of separating price from quantity in value added can be approached through the survey questions on changes in production volume. Information is available from each respondent on the per cent increase in output (in physical units) for the survey year. If we normalize price to equal 100 (say) for the survey year 1976, then the difference between the per cent increase in value added as computed above (e.g., (Q77-Q76)/Q76) and the per cent increase in production volume gives the per cent increase in price. It seems reasonable to take 1976 as the base year, even though this involves chaining both backwards and forwards in time, because the data for this year seem sounder than the 1975 data. It will be a worthwhile check (not yet performed) to compare the sectoral inflation rates computed using this method with their official statistical counterparts. If these derived price data seem coherent, perhaps they will provide an empirical basis for specifying an intra-sectoral distribution of prices within the model. Currently no such distribution is specified.

A final problem to be solved is what to do about those cases in which A21 = SUM, so that γ cannot be computed. This is not an infrequent occurrence -- in 1976 γ could not be computed for 32% of the respondents. We have not yet settled on a technique for surmounting this difficulty, but there are two obvious approaches. The first is to simply relate γ to other observable characteristics of the respondents and then to estimate the missing γ's using this relationship. In particular, γ is not a scale free measurement and seems to decrease systematically in the data with increases in QLIM. In addition, it may vary systematically between sectors. The second approach is to construct a time series of γ's for each firm and estimate a naive rate of growth (or decline) to use for extrapolation. Actually -- as mentioned above -- we have hopes of doing better than this; namely, to be able to relate rates of change in γ to a distributed lag of past investments.

APPENDIX

In this appendix some illustrations of the numerical magnitudes involved are presented. These should help the reader get a better feel for the data. Since the capacity utilization data are the key to the numerical algorithms these will be presented for the years 1975-76. The amount of excess capacity -- and especially the amount of excess labor -- that exists in Swedish industry will probably come as a surprise. Finally, plots of the functions QFR(L) for both 1975 and 1976 will be presented for selected firms both to illustrate the functional form and to illustrate the shifting mechanism.

The capacity utilization data are presented in Table 1. The survey compilation divides Swedish industry into 5 sectors -- (1) raw materials processings, (2) intermediate goods, (3) investment goods, (4) consumption goods and (5) building materials. This sectoral division does not fully match the one used in MOSES, but it is not difficult to reclassify firms. Note that since responses come from individual production units it sometimes happens that a firm is represented by respondents in 2 or 3 sectors.

This table gives a picture that is quite different from the one that is usually painted in economic models. The conventional treatment of capital as the fixed factor and labor as the variable factor implies that a reduction in output away from capacity primarily takes the form of a movement along the short run production frontier towards the origin, so that almost all excess capacity takes the form of unutilized capital. But these data indicate the contrary -- that labor is the relatively more fixed factor for most firms. Of course, there are some institutional peculiarities in Sweden which induce a bias towards fixity in labor. The most important is the Åman laws, dating from 1974, which require a pre-notification period of up to 6 months prior to a layoff.[1]

---

[1] An application of MOSES that illustrates the possible consequences of eliminating the Åman Laws is given in Eliasson [1977].

Unfortunately, comparable data dividing excess capacity in this fashion are not available in other countries (so far as I can tell), so it is difficult to judge how typical the Swedish situation is.[1]

Figures 5A-E present plots of 1975 and 1976 short run production frontiers for a selected individual firm in each of the 5 sectors. The upper asymptotes of these functions (i.e., normal capacity in 1975 and 1976) are also presented. Output (in millions of 1976 Skr) is graphed on the vertical axis against employment (in man-years) on the horizontal axis. The key for reading these figures is: + = QFR75, * = QFR76, ∇ = QLIM75, Δ = QLIM76. The method of generating the plots was simply to use the values of QLIM and γ generated by the algorithm presented in Section IV to compute QFR(L) for a sequence of values of L. Value added in 1976 was generated directly from the 1976 data using the adjustments for raw materials purchases and inventory changes explained above, and value added in 1975 was generated using the data on per cent changes in production volume from 1975 to 1976.

The firms pictured were chosen to illustrate the type of graphical portrayal of productions possibilities that can be given for all firms in the data bank. Of course, these are not "completely representative" firms; in particular, they are obviously all firms for whom γ could be computed for both years. Nonetheless, the plots illustrate the considerable diversity in relationships between employment and potential output that can be represented by the simple parameterization that is used in MOSES. They also show considerable diversity in the shifts of the short run production frontiers -- cf., the raw materials processing case in which increased capacity is "bought" at a price of reduced efficiency for lower levels of employment. This wide variability across firms' production structures coexists with considerable parameter stability within the individual firms from year to year.

---

[1] Some indirect evidence for the US based on a capital utilization series is offered by Solow [1972], who tentatively concludes that "....labour is more nearly the fixed factor in the short run, and variations in output are reflected substantially in the changing intensity of use of existing plant and equipment." (p.324)

Table 1

1975 Capacity Utilization... 3 July 1978

RAW MATERIALS PROCESSING

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	12
5-10 PCT	9
10-15 PCT	5
15-20 PCT	7
20-25 PCT	2
≥25 PCT	2

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	27
5-10 PCT	8
10-15 PCT	1
15-20 PCT	1
20-25 PCT	0
≥25 PCT	0

INTERMEDIATE GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	19
5-10 PCT	25
10-15 PCT	11
15-20 PCT	2
20-25 PCT	6
≥25 PCT	3

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	37
5-10 PCT	18
10-15 PCT	10
15-20 PCT	1
20-25 PCT	0
≥25 PCT	0

INVESTMENT GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	27
5-10 PCT	17
10-15 PCT	7
15-20 PCT	3
20-25 PCT	2
≥25 PCT	1

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	23
5-10 PCT	15
10-15 PCT	10
15-20 PCT	7
20-25 PCT	1
≥25 PCT	2

CONSUMPTION GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	10
5-10 PCT	15
10-15 PCT	6
15-20 PCT	5
20-25 PCT	0
≥25 PCT	2

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	15
5-10 PCT	10
10-15 PCT	6
15-20 PCT	4
20-25 PCT	1
≥25 PCT	2

Cont.

BUILDING MATERIALS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	4
5-10 PCT	10
10-15 PCT	3
15-20 PCT	1
20-25 PCT	0
≥25 PCT	0

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	6
5-10 PCT	4
10-15 PCT	3
15-20 PCT	4
20-25 PCT	1
≥25 PCT	0

RAW MATERIALS PROCESSING

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	0
5-10 PCT	10
10-15 PCT	8
15-20 PCT	5
20-25 PCT	4
≥25 PCT	1

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	18
5-10 PCT	6
10-15 PCT	3
15-20 PCT	1
20-25 PCT	0
≥25 PCT	2

INTERMEDIATE GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	15
5-10 PCT	9
10-15 PCT	11
15-20 PCT	11
20-25 PCT	5
≥25 PCT	4

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	20
5-10 PCT	15
10-15 PCT	10

INVESTMENT GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	21
5-10 PCT	17
10-15 PCT	7
15-20 PCT	0
20-25 PCT	3
≥25 PCT	2

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	12
5-10 PCT	18
10-15 PCT	12
15-20 PCT	2
20-25 PCT	3
≥25 PCT	3

CONSUMPTION GOODS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	12
5-10 PCT	11
10-15 PCT	4
15-20 PCT	3
20-25 PCT	0
≥25 PCT	0

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	9
5-10 PCT	12
10-15 PCT	2
15-20 PCT	2
20-25 PCT	4
≥25 PCT	1

BUILDING MATERIALS

DISTRIBUTION OF LABOR REDUNDANCY	
0-5 PCT	6
5-10 PCT	4
10-15 PCT	1
15-20 PCT	4
20-25 PCT	0
≥25 PCT	0

DISTRIBUTION OF CAPITAL REDUNDANCY	
0-5 PCT	4
5-10 PCT	4
10-15 PCT	3
15-20 PCT	3
20-25 PCT	0
≥25 PCT	1

Figure 5A. Raw Materials Processing

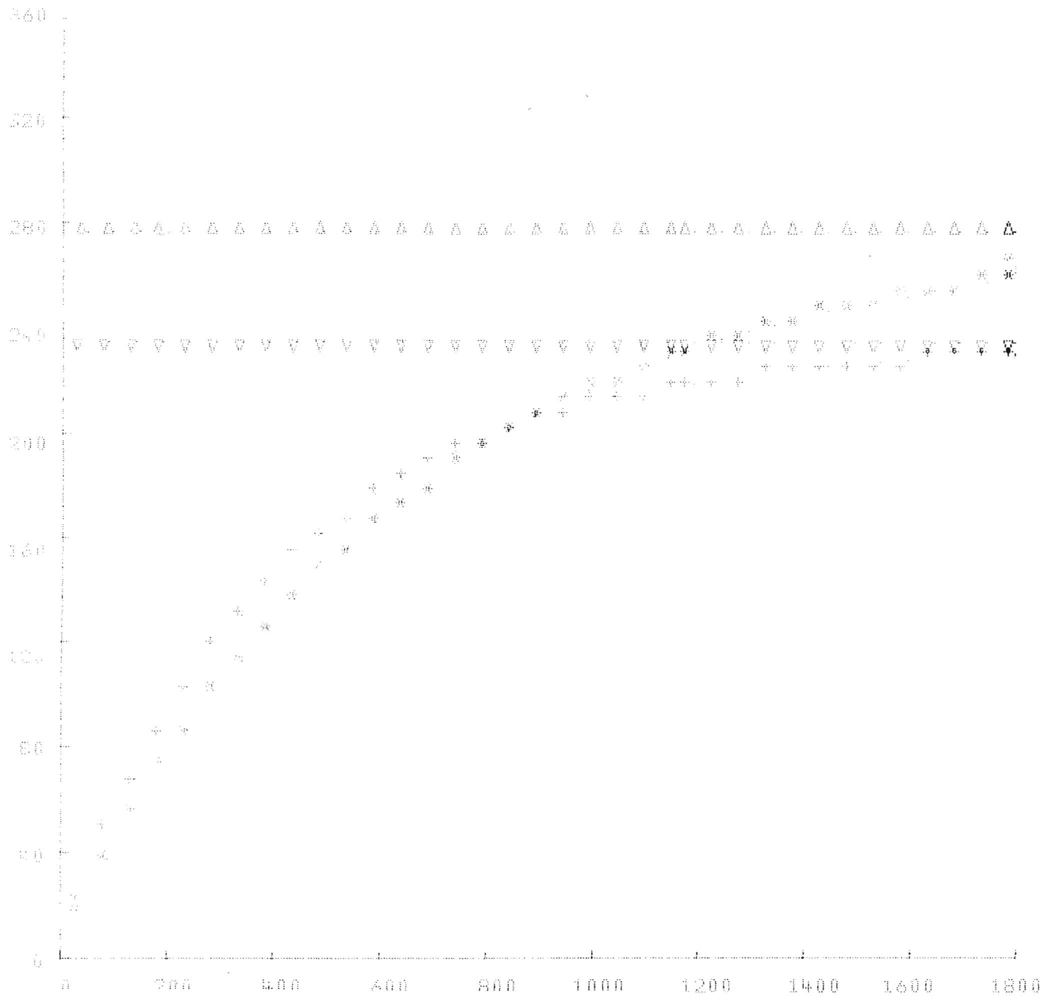


Figure 5B. Intermediate Goods

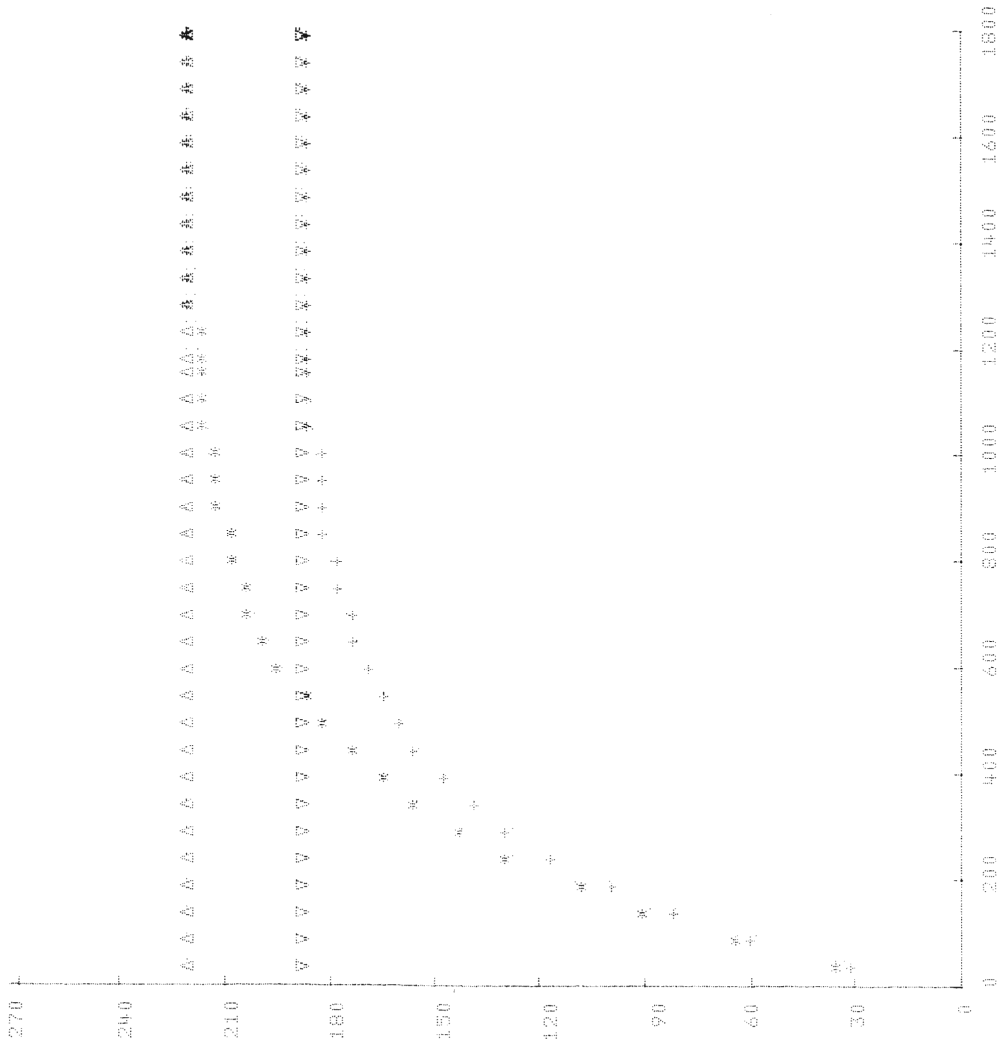


Figure [illegible]: Investment Goods

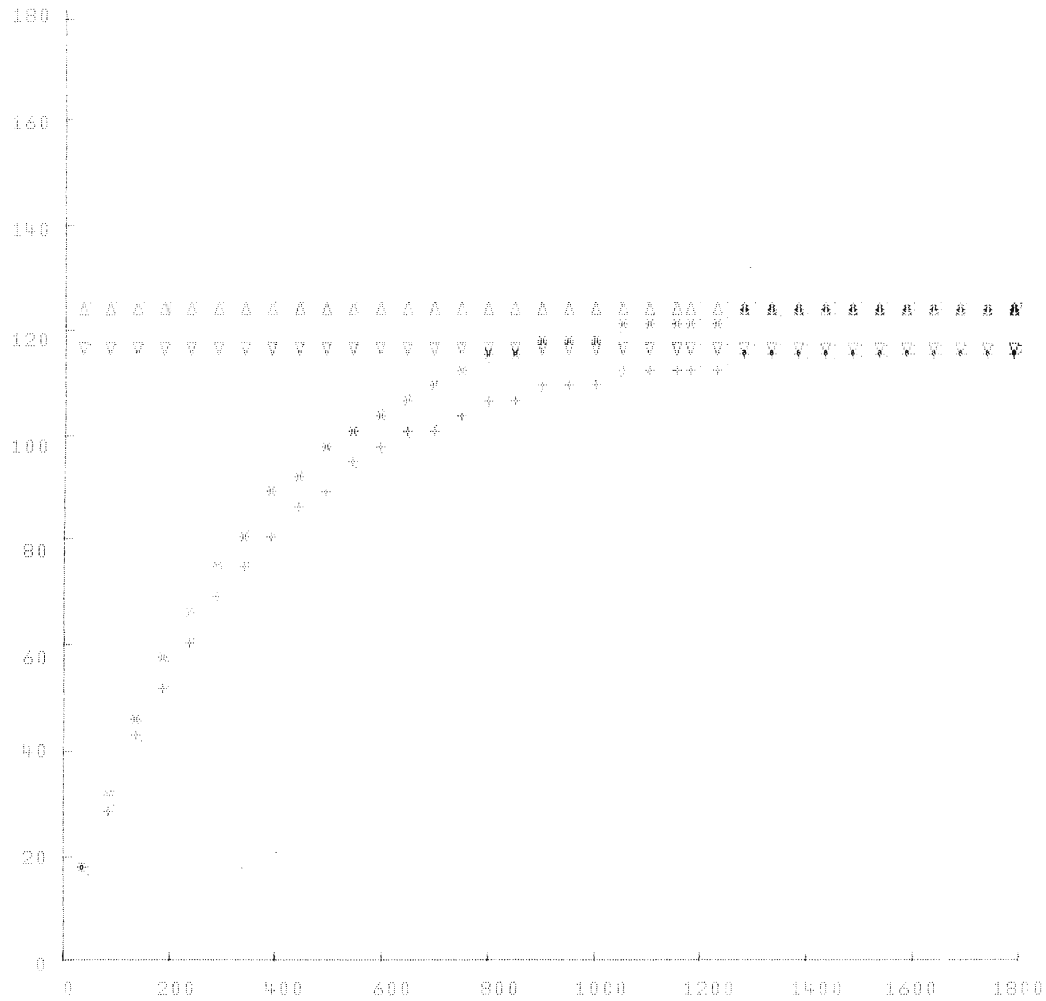


Figure 5D. Consumption Goods

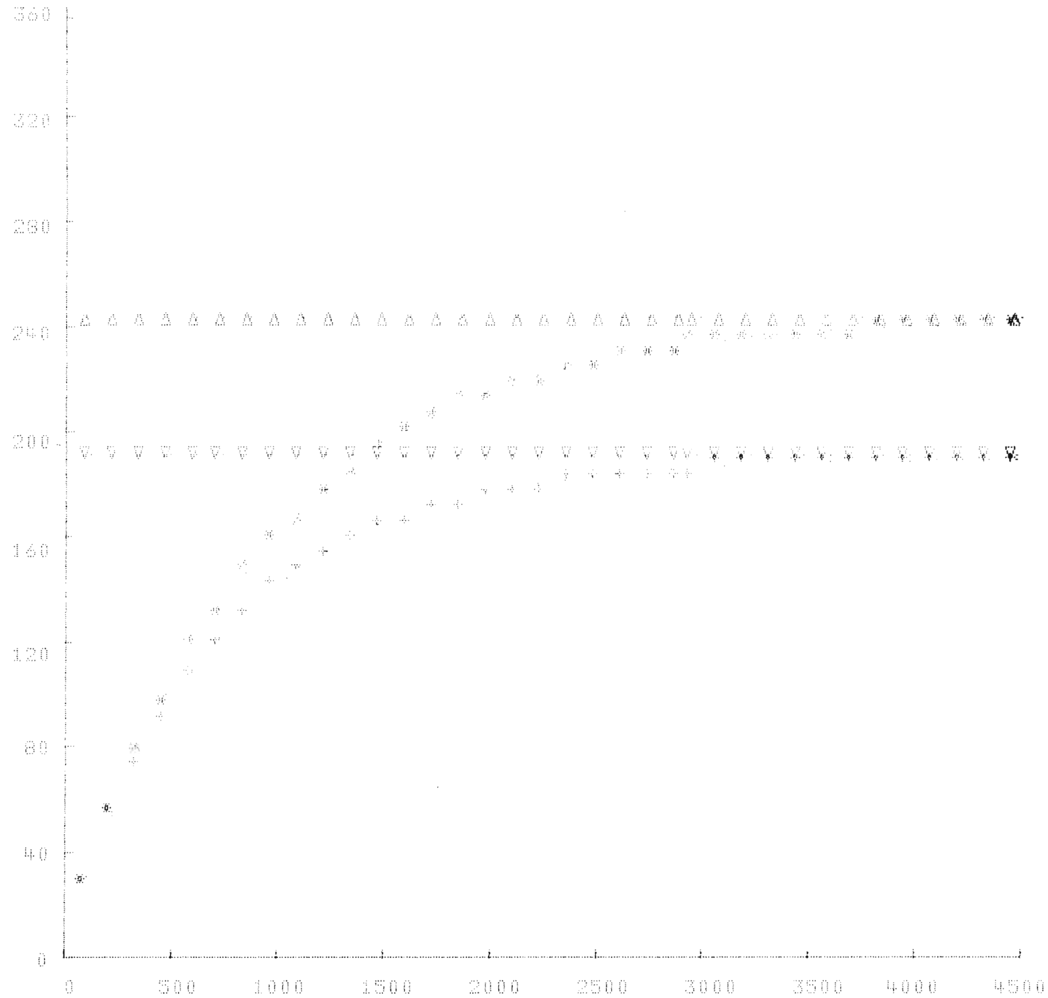
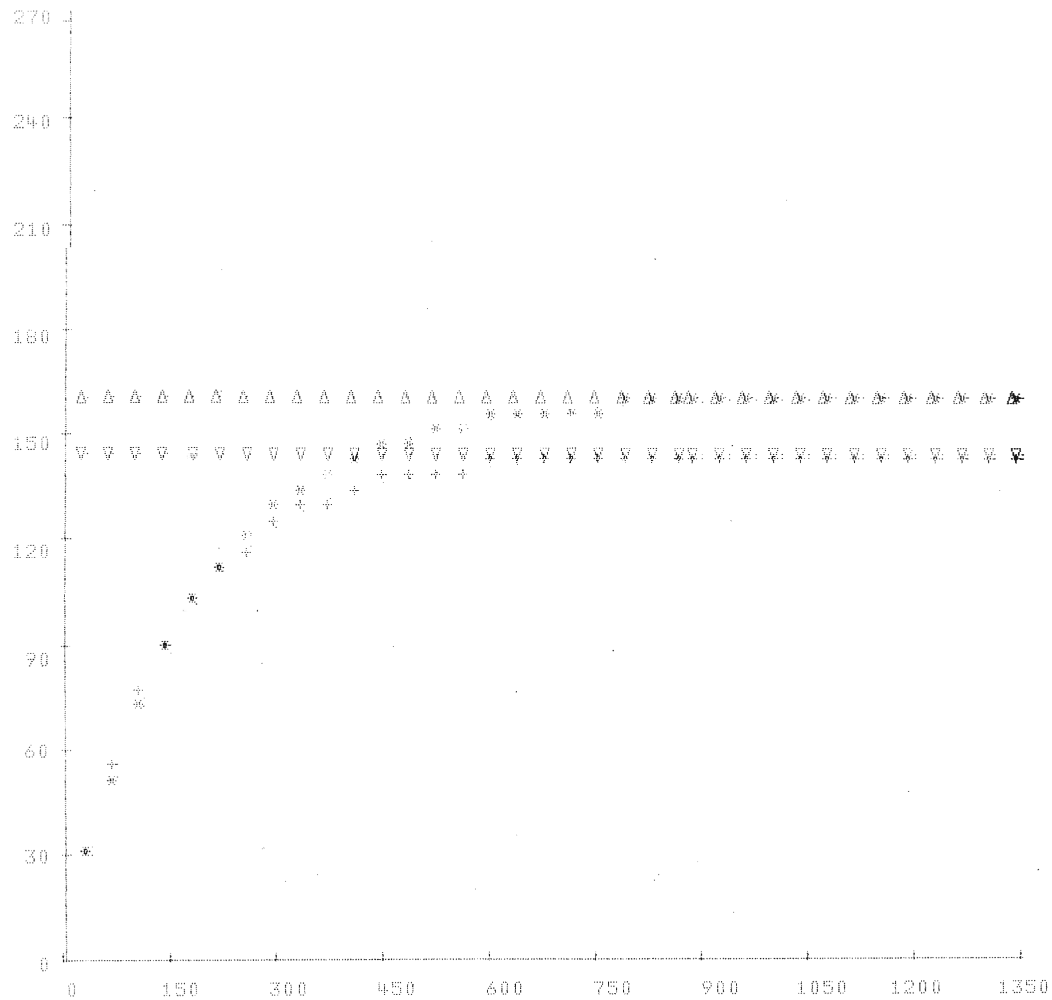


Figure 5E. Building Materials



REFERENCES

Dahlman, C J and Klevmarken, A, 1971, Den privata konsumtionen 1931-1975. Industriens Utredningsinstitut. Stockholm.

Eliasson, G, 1976, Business Economic Planning: Theory, Practice and Comparison. John Wiley & Sons. London.

— with the assistance of M Heiman and G Olavi, 1976, A Micro-to-Macro Interactive Simulation Model of the Swedish Economy. IUI Working paper No 7. Stockholm.

— , 1977, "Competition and Market Processes in a Simulation Model of the Swedish Economy". American Economic Review, Vol. 67, No. 1.

— , 1978, Experiments with Fiscal Policy Parameters on a Micro to Macro Model of the Swedish Economy; Paper prepared for the Conference on Micro Economic Simulation Models for the Analysis of Public Policy. The National Academy of Sciencies, Washington DC, March 1978.

(ed), 1978, A Micro-to-Macro Model of the Swedish Economy. IUI Conference Reports 1978:1. Stockholm.

Mayer, T, 1960, "Plant and equipment lead times", Journal of Business, 33, 1960, pp. 127-132.

Solow, R, 1972, "Some Evidence on the Short-Run Productivity Puzzle," Ch 17 in J Bhagwati and R Eckaus (eds), Development and Planning: Essays in Honor of Paul Rosenstein-Rodan. George Allen & Unwin. London.

Virin, O, 1976, "Industrins utveckling 1974-76 enligt Industriförbundets planenkät 1975/76," Specialstudie D, Industrikonjunkturen, Spring 1976. Sveriges Industriförbund. Stockholm.

A DYNAMIC FORMULATION OF THE LAW OF DIMINISHING RETURNS

Rolf Färe*
University of California, Berkeley

1. INTRODUCTION

The law of diminishing returns, originally formulated by Turgot [1844] for agriculture, has recently been investigated within a steady state framework for production (see Färe [1972], Färe & Jansson [1976], Shephard [1970a] and Shephard & Färe [1974]). In his work on the law Shephard [1970a] showed, for a single net output production technology, that for a bounded factor combination to limit output it is necessary and sufficient that it is essential By essentiality it is understood that only zero output is obtainable when the essential factors are null. He also showed that in general, not every positive bound on the essential factors leads to bounded output. On this issue Färe [1972][1] gave a sufficient condition. The work by Shephard on the law of diminishing returns was generalized in Shephard & Färe [1974] to hold for steady state multi-output production technologies.

A dynamic theory for production correspondences is being developed by Shephard and Färe, see Section 2 for details. Inputs and net outputs are treated as functions of time. For such production structures, two questions are important in relation to the law of diminishing returns. First, does there exist a positive bound on the rate (i.e., the norm) of an essential subvector of inputs such that the rate of net output is bounded? Second, does there exist a bound on the time availability (i.e.,the support) of an essential subvector of inputs such that net output is not available after a finite time horizon?

---
[1] Note that the condition given is not necessary as claimed.

* I am grateful to the Swedish Council for Social Science Research for sponsoring this research. I am also grateful to Professor R.W. Shephard for his important comments and suggestions.

This paper is addressed to each of the two questions and it is shown that there are bounds both on the rate and the time availability of essential inputs such that net output rate is bounded and such that net output is not available after a finite horizon. Consequently the smallest of these bounds serve to bound net output rate and net output (time) availability. The law of diminishing returns is understood to mean the existence of these bounds.

Necessary and sufficient conditions on the production structure, beyond the axioms, such that each bound on an essential factor combination bounds output rate and output (time) availability, are also given. In particular an input homothetic production structure satisfies these conditions.

2. THE GENERAL TECHNOLOGY

A general dynamic production structure is modelled here as in Shephard & Färe [1975] by an output correspondence $x \to \mathbb{P}(x)$ of input (vector) histories $x \in BM_+^n$ [1] to subsets of output histories $u \in BM_+^1$ (i.e., BM_+ for short) or inversely by an input correspondence $u \to \mathbb{L}(u)$ of output histories u to subsets of input (vector) histories x. $\mathbb{P}(x)$ denotes the set of all output histories obtainable from a vector of input histories $x \in BM_+^n$ and $\mathbb{L}(u)$ all input histories yielding at least the output history $u \in BM_+$. The two correspondences are inversely related by

$$\mathbb{L}(u) := \left\{x \in BM_+^n \mid u \in \mathbb{P}(x)\right\} \quad \text{and} \quad \mathbb{P}(x) := \{u \in BM_+ \mid x \in \mathbb{L}(u)\}.$$

$u \in BM_+$ denotes a net output history with $u(t)$ equal the number of units per unit time at time $t \in [0,+\infty)$. Similar for each $i \in \{1,2,\ldots,n\}$, $(x_1,x_2,\ldots,x_i,\ldots,x_n) \in BM_+^n$, $x_i(t)$ denotes the number of units per unit time at $t \in [0,+\infty)$.

The basic axioms taken for the dynamic production structure follow essentially those of Shephard & Färe [1975] and they are:

P.1 $\mathbb{P}(0) = 0$;

P.2 $\mathbb{P}(x)$ is bounded for $||x||$ finite;

P.3 $\mathbb{P}(\lambda \cdot x) \supset \mathbb{P}(x)$ for $\lambda \geqq 1$, $x \in BM_+^n$;

P.4 If $u \in \mathbb{P}(\lambda \cdot x)$ for some $\lambda > 0$ and $x \in BM_+^n$, then for each scalar $\theta \in (0,+\infty)$, there is a λ_θ such that $(\theta \cdot u) \in \mathbb{P}(\lambda_\theta \cdot x)$;

P.5 The correspondence $x \to \mathbb{P}(x)$ is closed (i.e., $[\{x_n\} \to x_0, \{u_n\} \to u_0$ with $u_n \in \mathbb{P}(x_n)$ for all $n] \Rightarrow u_0 \in \mathbb{P}(x_0)$);

---

[1] $BM_+^\alpha := \{f \in BM^\alpha \mid f(t) \geqq 0,\ t \in [0,+\infty)\}$, $\alpha = n$ or 1, where $BM^\alpha := \{f := (f_1,f_2,\ldots,f_\alpha) \mid f_i : [0,+\infty) \to \mathbb{R}$, f_i is bounded and measurable with $||f_i|| := \sup\{|f_i(t)| \mid t \in [0,+\infty)\}$ and the Euclidean product norm$\}$. BM^α is a Banach space, i.e., complete normed linear (see Shephard & Färe [1975]).

$\mathbb{P}.6 \quad u \in \mathbb{P}(x) \Rightarrow \{ v \mid 0 \leqq ||v|| \leqq ||u|| \} \subset \mathbb{P}(x),\ x \in BM_+^n;$

$\mathbb{P}$.A.A. The Asymmetric Axiom: The efficient subset of input histories, $\mathbb{E}ff\, \mathbb{L}(u) := \left\{ x \in BM_+^n \mid x \in \mathbb{L}(u),\ y \leq x\ ^{1)} \Rightarrow y \notin \mathbb{L}(u) \right\}$, is totally bounded [2] for $u \in BM_+^n$, $u \neq 0$, and $\mathbb{E}ff\, \mathbb{L}(0) := 0$.

The first property of the output correspondence $x \to \mathbb{P}(x)$ states that for null inputs, there can be only null output a self-evident axiom. For bounded input histories only bounded output is obtainable i.e., property $\mathbb{P}.2$ excludes e.g., the possibility of infinite accumulation. The third axiom is a statement concerning disposability of inputs. It says that individual input (vector) histories are disposable. Axiom four models obtainability and is motivated by the possibility of "doubling size of operation." The closeness axiom ($\mathbb{P}.5$) guarantees that there are efficient input and output histories. Note that $\mathbb{P}.5$ implies and is implied by that the input correspondence $u \to \mathbb{L}(u)$ is also closed. Disposability of output histories is modelled by axiom $\mathbb{P}.6$. This disposability axiom may be weakened to read $u \in \mathbb{P}(x)$ implies $\{v \mid v = \theta \cdot u,\ \theta \in [0,1]\} \subset \mathbb{P}(x)$, but for pedagogical reasons the stronger form $\mathbb{P}.6$ is applied in this paper. Finally, the asymmetric axiom is used to put a limit to input histories to be termed efficient.

Throughout this paper the above axioms are used as the basic model for production. Although frequently in the sequel the equivalent axioms on the input correspondence are used, they are all easy to derive and thus they are not listed here.

If only constant input and output histories are considered, the subspaces so obtained from BM^n and BM are isometrically isomorphic (i.e., equivalent) to $\mathbb{R}^n$ and $\mathbb{R}$, respectively. Consequently the steady state models discussed in Shephard [1970a,b] are special cases of the dynamic production structure.

---

1 $y \leq x$ means $y_i \leqq x_i$ $i = 1,2,\dots,n$ and $y_i < x_i$ for some i. $y_i < x_i$ means $y_i(t) \leq x_i(t)$ for all $t \in [0,+\infty)$ and $y_i(t) < x_i(t)$ for some $t \in [0,+\infty)$.

2 A set in BM_+^n is totally bounded if and only if every infinite sequence in the set contains a Cauchy subsequence.

3. PROPERTIES OF THE GENERAL TECHNOLOGY

A major interest in economic theory is to find efficient allocations. It is therefore important to show that the above production technology guarantees the existence of efficient input and output histories.

Proposition 1:

$\mathbb{L}(u)$ nonempty implies $\mathbb{E}\text{ff}\,\mathbb{L}(u)$ nonempty, $u \in \mathbb{BM}_+^n$.

Proof:

Assume $\mathbb{L}(u)$ nonempty for some $u \neq 0$ ($\mathbb{E}\text{ff}\,\mathbb{L}(0) := 0$) and let $x^o \in \mathbb{L}(u)$. Define $F_0 := \left(\mathbb{L}(u) \cap \left\{x \in \mathbb{BM}_+^n \mid x \leqq x^o\right\}\right)$ F_0 is a closed set as the intersection of two closed set. Furthermore, define $f(x^o) := \sup\{||x^o - x|| \mid x \in F_0\}$. $f(x^o) < +\infty$ is the diameter of F_0 measured from x^o. If $f(x^o) = 0$, then $x^o \in \mathbb{E}\text{ff}\,\mathbb{L}(u)$ and the proof is done. Thus assume $f(x^o) > 0$ and let x^* be an input vector such that $f(x^o) = ||x^o - x^*||$. Define $\bar{x} := (x^o - x^*)$ and consider $F_1 := \left(\mathbb{L}(u) \cap \left\{x \in \mathbb{BM}_+^n \mid x \leqq x^* + \bar{x}/2\right\}\right)$. Clearly $F_1 \subset F_0$ and F_1 is closed. By repeating this procedure one obtains; $F_n := \left(\mathbb{L}(u) \cap \left\{x \in \mathbb{BM}_+^n \mid x \leqq x^* + \bar{x}/2^n\right\}\right)$, $F_n \subset F_{n+1}$, $n = 1,2,\ldots$, F_n is closed and the diameter of F_n (i.e., $d(F_n) := \sup\{||x - y|| \mid x,y \in F_n\}$) tends to zero as $n \to \infty$.

It now follows from Cantor's Intersection Theorem, (see Simmons [1963], p.73), that $\left(\bigcap_{n=1}^{\infty} F_n\right) = \{x^*\}$ is a singleton. Consequently $x^* \in \mathbb{E}\text{ff}\,\mathbb{L}(u)$ and the proposition holds.

Q.E.D.

It is useful for the sequel to show that the input set $\mathbb{L}(u)$ is contained in a decomposition of input histories into those that are efficient and those that belong to $\mathbb{BM}_+^n$.

Proposition 2:

$$\mathbb{L}(u) \subset \left(\overline{\mathbb{E}ff\,\mathbb{L}(u)} + \mathbb{BM}_+^n \right), \quad ^{1} \quad u \in \mathbb{BM}_+ \; .$$

Proof:

If $u = 0$ or $u \neq 0$ with $\mathbb{L}(u)$ not empty the proposition holds. Thus for $x \in \mathbb{L}(u)$, $u \neq 0$ define the set $D(x) := \left\{ y \in \mathbb{BM}_+^n \mid y \leqq x \right\}$. From the proof of Proposition 1 it is clear that the intersection $\left(D(x) \cap \overline{\mathbb{E}ff\,\mathbb{L}(u)} \right)$ is nonempty. It is compact since $\overline{\mathbb{E}ff\,\mathbb{L}(u)}$ is compact and $D(x)$ is closed. Therefore, there is an input vector $z^* \in \left(D(x) \cap \overline{\mathbb{E}ff\,\mathbb{L}(u)} \right)$ such that $||z^*|| = \min \left\{ ||z|| \mid z \in \left(D(x) \cap \overline{\mathbb{E}ff\,\mathbb{L}(u)} \right) \right\}$ and hence, $x = z^* + (x - z^*)$ with $(x - z^*) \in \mathbb{BM}_+^n$ proving the proposition.

Q.E.D.

Frequently in economics, like in the theory of exhaustible resources (see Symposium on Exhaustible Resources [1974]) dynamic neoclassical production functions are applied. It is therefore of interest to determine their existence, hence introduce:

Definition 1:

The function $\Phi : \mathbb{BM}_+^n \to \mathbb{BM}_+$ defined pointwise by $\Phi(x;t) :=$ $\max \{u(t) \in \mathbb{R}_+ \mid u \in \mathbb{P}(x)\}$,[2] $t \in [0,+\infty)$, is called a dynamic neoclassical production function.

---

[1] For a set S, $\bar{S}$ denotes its closure.

[2] $\Phi(x;t)$ denotes the evaluation of $\Phi(x)$ at t.

Proposition 3:

There exists a dynamic neoclassical production function $\phi\phi(x) \varepsilon \mathbb{P}(x)$, $x \varepsilon BM_+^n$, if and only if the efficient subset of output histories $(\mathbb{E}\text{ff}\,\mathbb{P}(x))$ is a single output history.

Before proving this proposition define the efficient subset of output histories by:

$$\mathbb{E}\text{ff}\,\mathbb{P}(x) : = \begin{cases} \{u \mid u \varepsilon \mathbb{P}(x) \ , \ v \geq u \Rightarrow v \notin \mathbb{P}(x)\} \ , & \mathbb{P}(x) \neq 0 \\ 0 \ \text{ for } \ \mathbb{P}(x) = 0 \ . \end{cases}$$

The output correspondence $x \rightarrow \mathbb{P}(x)$ is bounded and closed (see properties $\mathbb{P}.2$ and $\mathbb{P}.5$) thus by argument like those of Proposition 1, it follows that $\mathbb{E}\text{ff}\,\mathbb{P}(x)$ nonempty for $x \varepsilon BM_+^n$.

Proof of Proposition 3:

Assume there is a dynamic neoclassical production function $\phi\phi(x) \varepsilon \mathbb{P}(x)$, $x \varepsilon BM_+^n$. Then $\phi\phi(x;t) \geqq u(t)$ for all $t \varepsilon [0,+\infty)$ and $u \varepsilon \mathbb{P}(x)$, implying that $\mathbb{E}\text{ff}\,\mathbb{P}(x) = \{\phi\phi(x)\}$. Conversely assume $\mathbb{E}\text{ff}\,\mathbb{P}(x) = \{u\}$ thus for all $v \varepsilon \mathbb{P}(x)$, $u(t) \geqq v(t)$, $t \varepsilon [0,+\infty)$ and hence $\phi\phi(x) : = u$ is a neoclassical dynamic production function.

Q.E.D.

4. ESSENTIALITY OF PRODUCTION FACTORS AND LIMITATIONALITET OF OUTPUT RATES

The first step in characterizing a dynamic law of diminishing returns is dealt with in this section. The aim here is to find conditions under which there are bounds on the rates of a subvector of input histories such that output rates are bounded even when the other inputs may freely vary. To pursue this issue introduce:

Definition 2:

A factor combination $\{\nu_1,\nu_2, \dots, \nu_k\}$, $1 \leqq k < n$, is essential if $\mathbb{P}(x) = 0$ for all $x \in \left\{x \in BM_+^n \mid x_{\nu_i} = 0, \quad i = 1,2, \dots, k\right\} =: D(\nu_1,\nu_2, \dots, \nu_k)$.

Definition 3:

A factor combination $\{\nu_1,\nu_2, \dots, \nu_k\}$, $1 \leqq k < n$, is output rate weak limitational if there exists a positive scalar B such that $\mathbb{P}(x)$ is bounded for all $x \in \left\{x \in BM_+^n \mid \|x_{\nu_1},x_{\nu_2}, \dots, x_{\nu_k}\| \leqq B\right\}$.

Definition 4:

A factor combination $\{\nu_1,\nu_2, \dots, \nu_k\}$, $1 \leqq k < n$, is output rate strong limitational if for each positive scalar B, $\mathbb{P}(x)$ is bounded for all $x \in \left\{x \in BM_+^n \mid \|x_{\nu_1},x_{\nu_2}, \dots, x_{\nu_k}\| \leqq B\right\}$.

Note that if a factor combination $\{\nu_1,\nu_2, \dots, \nu_k\}$ is essential, then the intersection $(\mathbb{L}(u) \cap D(\nu_1,\nu_2, \dots, \nu_k))$ is empty for all $u \neq 0$. Also note that an output rate strong limitational factor combination is weak limitational.

The relationship between essentiality and weak limitationality is clear from:

Proposition 4:

A factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is essential if and only if it is output rate weak limitational.

Proof:

Assume first that the factor combination $\{\nu_1,\nu_2, \ldots,\nu_k\}$ is not essential, then there is an input history $x^o \in D(\nu_1,\nu_2, \ldots, \nu_k)$ such that there is a nonzero output history $u \in \mathbb{P}(x^o)$, hence by property $\mathbb{P}$.4, that factor combination is not output rate strong nor weak limitational.

To prove the converse, assume that $\{\nu_1,\nu_2, \ldots, \nu_k\}$ is an essential factor combination. Then for any nonzero $u \in BM_+$, $\mathbb{L}(u) \cap D(\nu_1,\nu_2, \ldots, \nu_k)$ is empty. Also since $\mathbb{E}\text{ff}\,\mathbb{L}(u) \subset \mathbb{L}(u)$ and $\mathbb{L}(u)$ is closed (property $\mathbb{P}$.5) the intersection $\overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)} \cap D(\nu_1,\nu_2, \ldots, \nu_k)$, where $\overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)}$ denotes the closure of $\mathbb{E}\text{ff}\,\mathbb{L}(u)$, is empty. The set $D(\nu_1,\nu_2, \ldots, \nu_k)$ is nonempty and closed thus for $x \in \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}$, with $u^o \in BM_+$, $u^o \neq 0$ and $\mathbb{L}(u^o)$ nonempty, the distance

$$d(x,D(\nu_1,\nu_2, \ldots, \nu_k)) := \inf \{||x - y|| \mid y \in D(\nu_1,\nu_2, \ldots, \nu_k)\}$$

is strictly positive. The function d is continuous in x (see Berge [1963], p.84) and since $\overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}$ is a nonempty compact set (see asymmetric axiom) there is an input vector $x^o \in \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}$ such that x^o minimize

$$0 < \delta := \min \left\{d(x,D(\nu_1,\nu_2, \ldots, \nu_k)) \mid x \in \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}\right\}.$$

Choose as the positive bound $B = \delta/2$. By Proposition 2, $\mathbb{L}(u^o) \subset \left(\overline{\mathbb{E}\text{ff } \mathbb{L}(u^o)} + BM_+^n\right)$ thus it follows from property IP.6 that the intersections $\left(\mathbb{L}(u) \cap \left\{x \in BM_+^n \mid \|x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}\| \leqq B\right\}\right)$ are empty for all $\|u\| \geqq \|u^o\|$. Consequently, the essential factor combination is output rate weak limitational.

Q.E.D.

In order to show that essentiality not necessarily implies output rate strong limitationality, consider the following dynamic production function:

$$\phi\big(x_1(t), x_2(t)\big) := \begin{cases} g(x_1)\big(x_1(t) + x_2(t)\big) & \text{for } x \in X, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{where } X := \left\{ x \in BM_+^2 \mid x_i(t) := \begin{cases} c_o^i \in \mathbb{R}_+, & \text{if } t \in [0,T) \\ c_1^i \in \mathbb{R}_+, & \text{if } t \in [T,+\infty) \end{cases}, \; T>0,\; i=1,2 \right\}.$$

and $g(x_1) := \begin{cases} 1 & \text{if } x_1(t) \geqq B > 0,\ t \in [0,T) \\ 0 & \text{otherwise} \end{cases}$. Clearly the first factor is essential and for positive bound less that B, i.e., $\|x_1\| < B$, it is limitational. On the other hand output is not rate bounded for $\|x_1\| \geqq B$, since $\|x_2\|$ may be choosen arbitrarily large. Hence, x_1 is output rate weak, but not strong limitational.

For the special case of a homothetic input correspondence it will be shown that an essential factor combination is output rate strong limitational. Therefore introduce:

Definition 5:

The dynamic input correspondence $u \to \mathbb{L}(u)$ is homothetic if $\mathbb{L}(u) := F(u) \cdot \mathbb{L}(1)$, where the functional $F : BM_+ \to \mathbb{R}_+$ satisfies:

F.1 $F(u) > 0$ for $u \neq 0$. F.2 $F(u)$ is finite for $||u|| < +\infty$ and $\mathbb{L}(u)$ not empty, $+\infty$ for $\mathbb{L}(u)$ empty. F.3 $F(u) \geqq F(u')$ for $u \geqq u'$. F.4 F is lower semi-continuous. F.5 $F(u) \to +\infty$ as $||u|| \to +\infty$ with $\mathbb{L}(1)$ being a fixed input set, closed and for $x \in \mathbb{L}(1)$, $\lambda \cdot x \in \mathbb{L}(1)$, $\lambda \geqq 1$.

Proposition 5:

If $u \to \mathbb{L}(u)$ is homothetic, a factor combination $\{\nu_1, \nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is essential if and only if it is output rate strong limitational.

Proof:

From the first part of the proof of Proposition 4 it is clear that (in general without homotheticity) output rate strong limitationality of a factor combination implies that it is essential. To prove the converse let B^o be an arbitrarily chosen bound on the essential factors of production. The clearly from property F.5 there is a u^o such that the intersection $\left(F(u) \cdot \mathbb{L}(1) \cap \left\{x \in BM_+^n \mid ||x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}|| \leqq B\right\}\right)$ is empty for any $||u|| \geqq ||u^o||$. Q.E.D.

It is also of interest to give a complete characterization of when essentiality of a factor combination is equivalent to output rate strong limitationality. The following proposition does this.

Proposition 6:

A necessary and sufficient condition for an essential factor combination $\{\nu_1, \nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, to be output rate strong limitational is that for each positive bound B , there is a $u(B) \in BM_+$ such that the intersection $\left(\overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)} \cap \left\{x \in BM_+^n \mid ||x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}|| \leqq B\right\}\right)$ is empty for $||u|| \geqq ||u(B)||$.

Proof:

From Proposition 2 and property $\mathbb{P}.6$ of the technology the sufficiency clearly follows. To prove the necessity, assume that an essential factor combination is output rate strong limitational, then for each positive bound B, there is a $u(B) \in BM_+$ such that $\left(\mathbb{L}(u) \cap \left\{x \in BM_+^n \mid ||x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}|| \leqq B\right\}\right)$ is empty for any $||u|| \geqq ||u(B)||$. Consequently since $\mathbb{E}\mathrm{ff}\,\mathbb{L}(u) \subset \mathbb{L}(u)$ and $\mathbb{L}(u)$ is closed the proposition holds.

Q.E.D.

The economic interpretation of the condition stated in Proposition 6 is that for efficient increase in production, more of the essential factors must be used.

5. A REFINEMENT OF THE GENERAL TECHNOLOGY

The next step in explaining a dynamic law of diminishing returns is to show how time bonds on the essential factors i.e., $x_{\nu_i}(t) = 0$ for $t \geqq T_i$ $i = 1,2,\ldots,k$ relates to time bounds on net output. The purpose of this is to model the situation where some (essential) factor is exhausted and to analyze its consequences on the availability of net outputs.

In order to pursue this issue some additional axioms will be introduced. For this reason consider the following notations.

For $f \varepsilon BM_+^\alpha$ and $F \subset BM_+^\alpha$ ($\alpha = 1$ or n in this paper), define:

$$\text{supp } f := \left\{\overline{t \ \varepsilon \ \mathbb{R}_+^\alpha \mid f_i(t) > 0, \ i = 1,\ldots,\alpha}\right\},$$

$$\text{supp } F := \left\{\overline{t \ \varepsilon \ \mathbb{R}_+^\alpha \mid f_i(t) > 0, \ i = 1,\ldots,\alpha, \ f \ \varepsilon \ F}\right\}$$

$$\text{sup supp } F := \left\{t \ \varepsilon \prod_{i=1}^{\alpha} (\overline{\mathbb{R}}_+ \cup \{-1\})_i \mid t_i := \sup \{\tau_i \mid \tau_i \ \varepsilon \text{ supp } f_i\}, \ i = 1,\ldots,\alpha, \ f \ \varepsilon \ F\right\}.$$ [1)]

Note that sup supp for $f_i = 0$ is taken equal to -1 and for supp f_i not bounded sup supp $f_i := +\infty$.

The following three axioms on the time structure of production, not found in Shephard & Färe [1975], are applied:

T.1 $\mathbb{P}(x) = 0$ for $x \in BM_+^n$ with $x_i(t) = 0$ for $t > 0$, $i = 1,2,\ldots,n$ and sup supp $\mathbb{P}(x)$ is positive and bounded for some $x \in BM_+^n$;

T.2 If $(\text{sup supp } \mathbb{P}(x)) = t$, $t \in (0,+\infty)$, for $x \in BM_+^n$, then for each $\tau \in (0,+\infty)$, there is an input vector $\tilde{x} \in BM_+^n$, with $||\tilde{x}_i|| = ||x_i||$, $i = 1,2,\ldots,n$, such that $(\text{sup supp } \mathbb{P}(\tilde{x})) \geqq (t + \tau)$;

---

1 $\overline{\mathbb{R}}_+ : \mathbb{R}_+ \cup \{+\infty\}$, the positively extended nonnegative real numbers.

T.3 For supp u bounded, supp $\mathbb{E}\text{ff}\,\mathbb{L}(u)$ is bounded.

T.4 If sup supp u' $\geqq$ sup supp u , then sup supp $\mathbb{L}(u')$
sup supp $\mathbb{L}(u') \subset$ sup supp $\mathbb{L}(u)$.

The first part of axiom T.1 states that however large input applied only at t = 0, there can be no net output. Its second part says that output can be produced during a finite horizon. Note that T.1 dominates $\mathbb{P}$.1 above. Axiom T.2 expresses the idea that if net output is produced until time t, the production horizon can be extended to $(t+\tau)$ for any τ by extended the use of input in time. The third axiom states that for bounded production periods, it can not be efficient to apply in-inputs indefinitely, and the last property says that nondecreasing production horizon requires inputs to be applied at least as long time.

With these additional axioms on the production technology, time bounds on inputs and output is next studied.

6. ESSENTIALITY OF PRODUCTION FACTORS AND LIMITATIONALITY OF OUTPUT TIME

Two forms of output time limitationality are distinguished between, namely:

Definition 6:

A factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is output time weak limitational if there is a positive time bound $T \in (0,+\infty)$, such that sup supp $\mathbb{P}(x)$ is bounded for all $x \in \left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\ t \in [T,+\infty)\ ,\ i = 1,2, \ldots, k\right\}$,

and

Definition 7:

A factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is output time strong limitational if for each positive time bound $T \in (0,+\infty)$, sup supp $\mathbb{P}(x)$ is bounded for all $x \in \left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\right.$ $\left. t \in [T,+\infty)\ ,\ i = 1,2, \ldots, k\right\}$.

Clearly if a factor combination is output time strong limitational it is weakly so. Next the relationship between essentiality and output time weak limitationality is shown.

Proposition 7:

A factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is essential if and only if it is output time weak limitational.

Proof:

Assume first that the factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$ is not essential. Then there is an input history $x^o \in D(\nu_1,\nu_2, \ldots, \nu_k)$ such that $\mathbb{P}(x^o) \neq 0$. If sup supp $\mathbb{P}(x^o)$ not bounded then there is nothing to prove thus assume sup supp $\mathbb{P}(x^o)$ bounded. Then it follows from property T.2 that $\{\nu_1, \nu_2, \ldots, \nu_k\}$ is not output time weak nor strong limitational, proving the second part.

Q.E.D.

In proving the converse the following lemma is useful.

Lemma 1:

If the intersection $\left(\overline{\text{sup supp } \overline{\mathbb{E}\text{ff}}\, \mathbb{L}(u)} \cap \text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\ t \geqq T,\ i = 1,2, \ldots, k\right\}\right)$ is empty so is $\left(\text{sup supp } \mathbb{L}(u) \cap \text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\ t \geqq T,\ i = 1,2, \ldots, k\right\}\right)$, $u \in BM_+$.

Proof:

It is first shown that sup supp $\mathbb{L}(u) \subset \left(\overline{\text{sup supp } \overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)} + \mathbb{R}_+^n}\right)$.
Thus let $t \in$ sup supp $\mathbb{L}(u)$, then $t =$ sup supp x for some $x \in \mathbb{L}(u)$ and by Proposition 2, $x = y + z$ where $y \in \mathbb{E}\text{ff}\, \mathbb{L}(u)$ and $z \in BM_+^n$. Let $t_y :=$ sup supp y, then since $\overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)} \subset \mathbb{L}(u)$, $t - t_y \geqq 0$ and consequently, $t = t_y + (t - t_y)$ where $t_y \in \overline{\text{sup supp } \overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)}}$ and $(t - t_y) \in \overline{\mathbb{R}}_+^n$. It is clear that supp $\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\ t \geqq T,\ i = 1,2, \ldots, k\right\} \subset$
$\left\{t \in \prod_{i=1}^{n} (\overline{\mathbb{R}}_+ \cup \{-1\})_i \mid t_{\nu_i} \leqq T,\ i = 1,2, \ldots, k,\ t_{\nu_i} \in \overline{\mathbb{R}}_+ \cup \{-1\},\ i = k+1, k+2, \ldots, n\right\}$ and hence the lemma holds.

Q.E.D.

To continue the proof of Proposition 7 it is next shown that intersection $\left(\overline{\text{sup supp } \overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)}} \cap \text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0,\ t > 0,\ i = 1,2, \ldots, k\right\}\right)$ is empty for $u \neq 0$. For this reason assume that t^o belongs to the intersection. Then there is a sequence $\{t^n\} \subset \overline{\text{sup supp } \overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)}}$ with $t^n \to t^o$ as $n \to +\infty$.

Consequently there is a sequence of input histories $\{x^n(t^n)\} \subset \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)}$. It then follows from the compactness of $\overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)}$ that $x^{n_\ell} \to x^o \in \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)}$ for some subsequence x^{n_ℓ} of $x^n(t^n)$. Hence by essentiality of the factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$ and by property T.1 of the production technology x^o must have $x_{\nu_i}(t) > 0$ for some $t > 0$, and $i = 1,2, \ldots, k$ a contradiction since t^o was picked from the intersection $\left(\sup \text{supp}\ \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u)} \cap \text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0\ ,\ t > 0\ ,\ i = 1,2, \ldots, k\right\}\right)$. Now let $u^o \in BM_+^n$, $u^o \neq 0$, supp u^o bounded and $\mathbb{L}(u^o)$ not empty. Then, $\sup \text{supp}\ \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}$ is a nonempty compact subset of $\prod_{i=1}^{n} (\overline{\mathbb{R}}_+ \cup \{-1\})_i$ (see property T.3), the set $\text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0\ ,\ t > 0\ ,\ i = 1,2, \ldots, n\right\}$ is nonempty and by definition closed, consequently by arguments like those used to prove Proposition 4, there is a positive time bound T such that $\sup \text{supp}\ \overline{\mathbb{E}\text{ff}\,\mathbb{L}(u^o)}$ has an empty intersection with $\text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0\ ,\ t \geqq T\ ,\ i = 1,2, \ldots, k\right\}$. Consequently by Lemma 1 and property T.4, Proposition 7 holds.

Q.E.D.

The production function (1) above also satisfies properties T.1 - T.4. The first factor is essential and for time bounds on x_1 different from T, it is output time limitational. However, for $x_1(t) \geqq B$, $t \in [0,T)$ and $x_1(t) = 0$, $t \in [T,+\infty)$, $x_2(t) > 0$ for $t \in [T,+\infty)$, $x = (x_1,x_2) \in X$, the support of output is not bound, consequently, x_1 is output time weak but not strong limitational.

For a homothetic input correspondence the following proposition is valid:

Proposition 8:

If the input correspondence $u \to \mathbb{L}(u)$ is homothetic,an essential factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, is output time strong limitational.

If for efficiently increased production time of net output (i.e., sup supp u), the use of the essential factors have to be extended in time, then a factor combination is output time strong limitational if and only if it is essential. Formally:

Proposition 9:

A necessary and sufficient condition for an essential factor combination $\{\nu_1,\nu_2, \ldots, \nu_k\}$, $1 \leqq k < n$, to be output time strong limitational is that for each positive time bound T there is an output history $u(T)$ such that the intersection $\left(\overline{\text{sup supp } \overline{\mathbb{E}\text{ff}\, \mathbb{L}(u)}} \cap \text{supp}\left\{x \in BM_+^n \mid x_{\nu_i}(t) = 0 \ , \ t \geqq T \ , \ i = 1,2, \ldots, k\right\}\right)$ is empty for sup supp $u \geqq$ sup supp $u(T)$.

The proof is immediate from Lemma 1 and property T.4 of the production structure.

REFERENCES

Berge, C., 1963, Topological Spaces, McMillan Company.

Färe, R., 1972, "Strong Limitationality of Essential Proper Subsets of Factors of Production," Zeitschrift für Nationalökonomie, Vol.34, pp.417-424.

Färe, R. and Jansson, L., 1976, "Joint Inputs and the Law of Diminishing Returns," Zeitschrift für Nationalökonomie, Vol.36, pp. 407-416.

Shephard, R.W., 1970a, "Proof of the Law of Diminishing Returns," Zeitschrift für Nationalökonomie, Vol.30, pp.7-34.

———, 1970b, Theory of Cost and Production Functions, Princeton University Press.

Shephard, R.W. and Färe, R., 1974, "The Law of Diminishing Returns," Zeitschrift für Nationalökonomie, Vol.34, pp.69-90.

Shephard, R.W. and Färe, R., 1975, "A Dynamic Theory of Production Correspondences," ORC 75-13, Operations Research Center, University of California, Berkeley, Calif.

Simmons, G.F., 1963, Introduction to Topology and Modern Analysis, McGraw-Hill Book Company, Inc.

Symposium on Exhaustible Resources, 1974, Review of Economic Studies. December 1974.

Turgot, A.R.J., 1844, "Observations sur le Memoire de M.Saint-Peravy 1667, "in Oeuves de Turgot, (ed. Daire), Vol.1, pp.418-483. Paris.

ON RAY-HOMOTHETIC PRODUCTION FUNCTIONS*

Rolf Färe
University of California
Berkeley

Leif Jansson
IUI
Stockholm

C.A. Knox Lovell
University of North
Carolina at Chapel Hill

1. Homogeneous scalar valued production functions have played an important role in econometric studies of production for estimating returns to scale. But since this class of functions models very simple technologies, others have been developed. Shephard [1953], [1970] introduced the class of homothetic functions, in which returns to scale can vary with output but not with the input mix. Eichhorn [1969], [1970] derived the class of ray-homogeneous functions by solving a multiplicative Cauchy functional equation. For such a class returns to scale can vary with the input mix, but not with output. The homothetic and ray-homogeneous classes were combined by Färe [1973], who solved a translation functional equation to obtain the class of ray-homothetic functions. Such functions are homothetic along each ray in input space, but possibly in different ways for different rays. As a result, returns to scale can vary both with output and with the input mix. It naturally follows that technically optimal (i.e., cost minimizing) output can vary both with output and with the input mix when the production function is ray-homothetic.

Homothetic production functions have been estimated by Zellner and Revankar [1966] among many others, but to the best of our knowledge neither ray-homogeneous nor ray-homothetic production functions have ever been estimated. The present paper represents an attempt to fill that gap by specifying and providing estimates of a ray-homothetic production function. We also demonstrate that the implications of ray-homotheticity for returns to scale, and

---

* Sponsored by Anders Otto Swärds stiftelse.

hence for technically optimal output, differ substantially from those of homotheticity and ray-homogeneity.

2. Let $\phi:R_+^n \to R_+$ be a production function with properties:[1]

ϕ.1 $\phi(0) = 0$, and $\phi(x) > 0$ for some $x \geq 0$.[2]

ϕ.2 ϕ is bounded for bounded input vectors x.

ϕ.3 $\phi(\lambda \cdot x) \geqq \phi(x)$ for $\lambda \geqq 1$.

ϕ.4 For any $x \geq 0$ such that $\phi(\lambda \cdot x) > 0$ for some scalar $\lambda > 0$, $\phi(\lambda \cdot x) \to +\infty$ as $\lambda \to +\infty$.

ϕ.5 ϕ is upper semi-continuous.

Also, consider the functions $F:R_+ \to R_+$ and $H: \{x/|x| \,|\, x \geq 0\} \to R_+$ with the properties

F.1 $F(0) = 0$.

F.2 $F(v)$ is bounded for $|v| < +\infty$.

F.3 F is strictly increasing.

F.4 $F(v) \to +\infty$ as $v \to +\infty$.

F.5 F is continuous.

H. $H(x/|x|) > 0$ and bounded.

A production function ϕ is ray-homothetic if

$$(1) \qquad \phi(\lambda \cdot x) = F(\lambda^{H(x/|x|)} \cdot G(x)), \quad \lambda > 0,$$

where $G(x) = F^{-1}(\phi(x))$. If F is the identity function, then

$$(2) \qquad \phi(\lambda \cdot x) = \lambda^{H(x/|x|)} \cdot \phi(x), \quad \lambda > 0,$$

and ϕ is ray-homogeneous. On the other hand, if $H(x/|x|)$ is a positive constant α, then

$$(3) \qquad \phi(\lambda \cdot x) = F(\lambda^{\alpha} \cdot G(x)), \quad \lambda > 0,$$

---

[1] These properties are adapted from Shephard [1974], who also assumes that the efficient subsets are bounded.

[2] $x \geq 0$ means $x \geqq 0$ but $x \neq 0$.

and ϕ is homothetic. Thus the ray-homothetic function (1) provides a straightforward generalization of the functions of Eichhorn and Shephard. Finally, if F is the identity function and $H(x/|x|)$ is a positive constant, then

$$(4) \quad \phi(\lambda \cdot x) = \lambda^{\alpha} \circ \phi(x), \quad \lambda > 0,$$

and ϕ is homogeneous.

Goldman and Shephard [1972] have proved that the ray-homogeneous function (2) satisfies (global) strong disposability of inputs ($x' \geqq x => \phi(x') \geqq \phi(x)$) or (global) quasiconcavity if and only if $H(x/|x|)$ is a positive constant, in which case it is homogeneous. Färe [1975] has proved a similar theorem stating that the ray-homothetic function (1) satisfies the same two (global) properties if and only if $H(x/|x|)$ is a positive constant, in which case it is homothetic. Although neither of these two strong properties is imposed globally by the ray-homothetic function, they may be satisfied locally (i.e., for some neighborhood of a point $x \varepsilon R^n_+$) even if $H(x/|x|)$ is not a positive constant.

Defining the elasticity of scale ε as

$$\varepsilon = \lim_{\lambda \to 1} \left[\frac{\partial \phi(\lambda \cdot x)}{\partial \lambda} \quad \frac{\lambda}{\phi(x)}\right],$$

one can easily calculate this elasticity, assuming sufficient regularity, for the above functions. Clearly $\varepsilon = \varepsilon_1(x/|x|, \phi(x))$ for the ray-homothetic function, $\varepsilon = \varepsilon_2(x/|x|)$ for the ray-homogeneous function, $\varepsilon = \varepsilon_3(\phi(x))$ for the homothetic function, and $\varepsilon = \varepsilon_4 = \alpha$, a constant, for the homogeneous function. Technically optimal output is obtained for the ray-homothetic and homothetic functions by setting $\varepsilon_i = 1$, $i = 1,3$. Technically optimal output is zero, indeterminate or infinite for the ray-homogeneous and homogeneous functions.

3. In their article on generalized production functions, Zellner and Revankar [1966] discuss various properties of homothetic production functions. They also provide an econometric example showing how a

parametric homothetic production function can be estimated. However a simple inspection of a plot of their data, along lines suggested by Hanoch and Rothschild [1972], led us to conclude that the data need not necessarily have been generated from a technology satisfying (global) strong disposability of inputs or (global) quasi concavity. For that reason we demonstrate how a parametric ray-homothetic production function can be estimated and interpreted. We borrow the data, and a portion of the parametric specification, from Zellner and Revankar. The functional specification is

$$(5) \quad Ve^{\theta V} = AK^{\alpha+\gamma(K/L+\delta L/K)^{-1}} L^{\beta+\gamma(K/L+\delta L/K)^{-1}},$$

with θ, $\gamma \varepsilon R$, A, α, β, $\delta \varepsilon R_+$, and $[\alpha + \gamma(K/L + \delta L/K)^{-1}] > 0$, $[\beta + \gamma(K/L + \delta L/K)^{-1}] > 0$, for all K/L. If $\gamma = 0$ then (5) is the homothetic Cobb-Douglas function used by Zellner and Revankar. If $\theta = 0$ then (5) is ray-homogeneous, and if $\theta = \gamma = 0$ then (5) is a homogeneous Cobb-Douglas function.

For the statistical model we follow the methodology of Zellner, Kmenta and Drèze [1969] by assuming that the data were generated by a process consistent with the maximization of the mathematical expectation of profits.

Introducing a multiplicative random error term in (5) and taking natural logarithms gives the estimating equation

$$(6) \quad \ln V_i + \theta V_i = \ln A + \left[\alpha + \gamma((K_i/L_i) + \delta(L_i/K_i))^{-1}\right] \ln K_i + \left[\beta + \gamma((K_i/L_i) + \delta(L_i/K_i))^{-1}\right] \ln L_i + \mu_i ,$$

where i = 1,...,25 indexes observations. The variables, V, K, L refer to per-establishment means of value added, capital and labor for each of 25 states in the U.S. Transportation Equipment Industry in 1957, and are described in greater detail by Zellner and Revankar. It is assumed that $\mu_i \sim NID(0,\sigma^2)$, and that $E(\mu_i \mu_j) = 0$, $i \neq j$. Under these assumptions the parameters of (6) may be estimated by maximum likelihood methods. The results are presented in Table 1;

column (1) contains estimates of the ray-homothetic function (5), while columns (2) and (3) contain estimates of the ray-homogeneous and homothetic versions of (5) respectively.

All three specifications provide excellent fits to the data, although the least restrictive of the three, the ray-homothetic function, is clearly to be preferred. Estimates of all parameters of the ray-homothetic function are highly significant. The estimated value of θ is significantly greater than zero, suggesting that technology is not ray-homogeneous; and the estimated value of γ is significantly less than zero, suggesting that technology is not homothetic either. The estimated ray-homothetic function is depicted by a series of isoquants in Figure 1. The single dashed isoquant belongs to the estimated homothetic function.

4. The empirical estimates obtained above can be used to draw some inferences for returns to scale and technically optimal output. Applying the definition of ε to the parametric ray-homothetic production function (5) gives

$$(7) \quad \varepsilon\left(x/|x|,\ \theta(x)\right) = \frac{\alpha + \beta}{1 + \theta V} + \frac{2\gamma}{(1+\theta V)(K/L+\delta L/K)}.$$

At technically optimal output, $\varepsilon_1(x/|x|, \theta(x)) = 1$, and thus

$$(8) \quad V_1^o = \frac{\alpha + \beta - 1}{\theta} + \frac{2\gamma}{\theta(K/L+\delta L/K)}.$$

Both $\varepsilon_1(x/|x|,\ \theta(x))$ and V_1^o can be computed for each observation, using parameter estimates given in Table 1. Computed values of $\varepsilon_1(x/|x|,\ \theta(x))$ measure returns to scale at each observation, while computed values of V_1^o can be compared with actual values of V for each observation to determine the magnitude of the resulting deviation of actual from technically optimal output. These results are given in Table 2, along with analogous results for the ray-homgeneous and homothetic versions of (5). Table 2 emphasizes shortcomings of the latter two functions that are not otherwise apparent. For the homothetic function returns to scale is a (monotonically decreasing) function of output only, and so technically optimal output is the same constant for all observations. For the ray-homogeneous function returns to scale is a U-shaped function of the input mix only, reaching a minimum at $K/L = \delta^{1/2} = 0.844$. Since this minimum value exceeds unity, technically optimal output is infinite for all observations. Neither of these scenarios is plausible.

For the ray-homothetic function, however, returns to scale is a monotonically decreasing function of output and a U-shaped function of the input mix, reaching a minimum with respect to the latter at $K/L = \delta^{1/2} = 0.635$. As a result, technically optimal output varies across observations, as one would expect. Despite this variation the majority of production is carried out in the region of increasing returns to scale and so actual output is on average only 65.5 % of technically optimal output.

5. The ray-homothetic function includes ray-homogeneous and homothetic functions as special cases, and is considerably more flexible than either. We have constructed and estimated a parametric version of a ray-homothetic function, using a Cobb-Douglas function as a base. Undoubtedly more complex bases can be used (e.g., the CES function), but there seems to be no reason to do so. Our specification is relatively easy to estimate, and it is sufficiently flexible to permit returns to scale to attain a different value at every point in input space. This flexibility of the elasticity of scale in turn permits technically optimal output to vary with the input mix, a desirable property that is absent in both the homothetic and the ray-homogeneous functions.

Table 1. Estimated Production Function Parameters

Parameter	Ray-Homothetic (1)	Ray-Homogeneous (2)	Homothetic (3)
Θ	0.098 (0.009)	-	0.114 (0.015)
A	14.941 (1.007)	7.989 (1.075)	19.298 (1.891)
α	0.330 (0.026)	0.221 (0.083)	0.355 (0.026)
β	1.440 (0.046)	1.319 (0.186)	1.104 (0.380)
γ	-0.259 (0.047)	-0.403 (0.351)	-
δ	0.403 (0.096)	0.712 (0.774)	-
$\bar{R}^2$	0.969	0.957	0.919
$\ln\alpha$	6.526	6.323	5.633

Figures in parentheses are asymptotic standard errors.

Table 2. Implied Values of Returns to Scale and Technically Optimal Output

State	V	K/L	Ray-Homothetic $\varepsilon_1(x/\|x\|,\phi(x))$	Ray-Homothetic V_1^o	Ray-Homogeneous $\varepsilon_2(x/\|x\|)$	Ray-Homogeneous V_2^o	Homothetic $\varepsilon_3(\phi(x))$	Homothetic V_3^o
Florida	0.193	0.341	1.403	4.387	1.208	$+\infty$	1.428	4.026
Maine	0.364	0.304	1.402	4.617	1.215	"	1.401	"
Iowa	0.477	0.337	1.368	4.415	1.211	"	1.383	"
Louisiana	0.638	0.237	1.414	5.135	1.291	"	1.340	"
Massachusetts	1.404	0.389	1.236	4.149	1.176	"	1.258	"
West Virginia	1.513	0.380	1.228	4.192	1.182	"	1.244	"
Texas	1.712	0.207	1.310	5.405	1.318	"	1.221	"
Alabama	1.855	0.121	1.371	6.335	1.405	"	1.204	"
New York	2.040	0.384	1.174	4.173	1.179	"	1.184	"
Virginia	2.052	0.229	1.257	5.202	1.298	"	1.182	"
California	2.333	0.410	1.138	4.067	1.164	"	1.152	"
Wisconsin	2.463	0.417	1.124	4.041	1.160	"	1.139	"
Illinois	2.629	0.667	1.084	3.702	1.075	"	1.122	"
Pennsylvania	2.651	0.350	1.131	4.341	1.201	"	1.121	"
New Jersey	2.701	0.401	1.108	4.101	1.169	"	1.115	"
Maryland	3.219	0.253	1.132	4.997	1.277	"	1.067	"
Washington	3.558	0.350	1.057	4.339	1.201	"	1.038	"
Indiana	3.816	0.760	0.996	3.764	1.065	"	1.017	"
Kentucky	4.031	0.434	0.996	3.979	1.151	"	1.000	"
Georgia	4.289	0.253	1.049	4.995	1.277	"	0.980	"
Ohio	4.440	0.608	0.950	3.701	1.086	"	0.969	"
Connecticut	4.485	0.320	1.002	4.513	1.223	"	0.966	"
Missouri	5.217	0.387	0.931	4.159	1.178	"	0.915	"
Kansas	6.507	0.120	0.990	6.339	1.406	"	0.838	"
Michigan	7.182	0.887	0.813	3.920	1.063	"	0.802	"

Table 1. Estimated Production Function Parameters

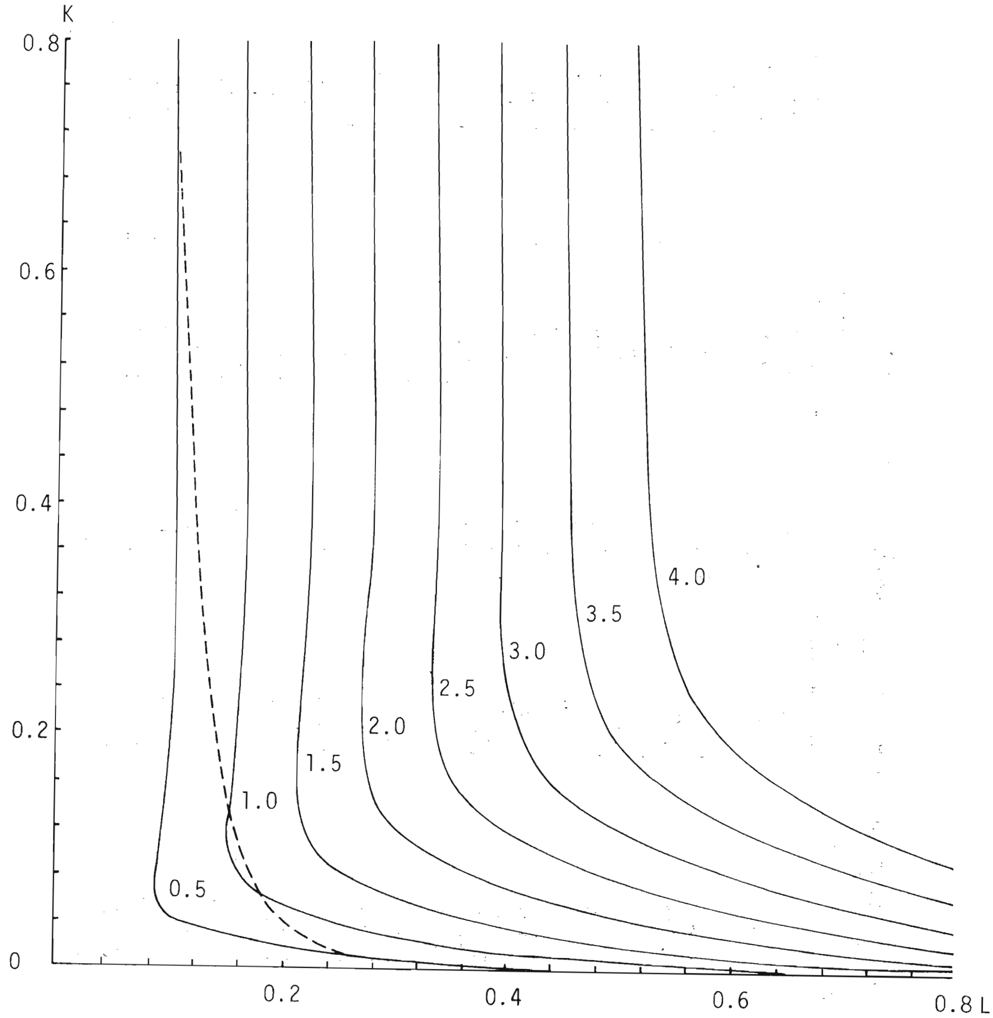
Parameter	Ray-Homothetic (1)	Ray-Homogeneous (2)	Homothetic (3)
Θ	0.098 (0.009)	-	0.114 (0.015)
A	14.941 (1.007)	7.989 (1.075)	19.298 (1.891)
α	0.330 (0.026)	0.221 (0.083)	0.355 (0.026)
β	1.440 (0.046)	1.319 (0.186)	1.104 (0.380)
γ	-0.259 (0.047)	-0.403 (0.351)	-
δ	0.403 (0.096)	0.712 (0.774)	-
$\bar{R}^2$	0.969	0.957	0.919
$\ln\alpha$	6.526	6.323	5.633

Figures in parentheses are asymptotic standard errors.

Table 2. Implied Values of Returns to Scale and Technically Optimal Output

			Ray-Homothetic		Ray-Homogeneous		Homothetic	
State	V	K/L	$\varepsilon_1(x/\|x\|,\phi(x))$	V_1^o	$\varepsilon_2(x/\|x\|)$	V_2^o	$\varepsilon_3(\phi(x))$	V_3^o
Florida	0.193	0.341	1.403	4.387	1.208	$+\infty$	1.428	4.026
Maine	0.364	0.304	1.402	4.617	1.215	"	1.401	"
Iowa	0.477	0.337	1.368	4.415	1.211	"	1.383	"
Louisiana	0.638	0.237	1.414	5.135	1.291	"	1.340	"
Massachusetts	1.404	0.389	1.236	4.149	1.176	"	1.258	"
West Virginia	1.513	0.380	1.228	4.192	1.182	"	1.244	"
Texas	1.712	0.207	1.310	5.405	1.318	"	1.221	"
Alabama	1.855	0.121	1.371	6.335	1.405	"	1.204	"
New York	2.040	0.384	1.174	4.173	1.179	"	1.184	"
Virginia	2.052	0.229	1.257	5.202	1.298	"	1.182	"
California	2.333	0.410	1.138	4.067	1.164	"	1.152	"
Wisconsin	2.463	0.417	1.124	4.041	1.160	"	1.139	"
Illinois	2.629	0.667	1.084	3.702	1.075	"	1.122	"
Pennsylvania	2.651	0.350	1.131	4.341	1.201	"	1.121	"
New Jersey	2.701	0.401	1.108	4.101	1.169	"	1.115	"
Maryland	3.219	0.253	1.132	4.997	1.277	"	1.067	"
Washington	3.558	0.350	1.057	4.339	1.201	"	1.038	"
Indiana	3.816	0.760	0.996	3.764	1.065	"	1.017	"
Kentucky	4.031	0.434	0.996	3.979	1.151	"	1.000	"
Georgia	4.289	0.253	1.049	4.995	1.277	"	0.980	"
Ohio	4.440	0.608	0.950	3.701	1.086	"	0.969	"
Connecticut	4.485	0.320	1.002	4.513	1.223	"	0.966	"
Missouri	5.217	0.387	0.931	4.159	1.178	"	0.915	"
Kansas	6.507	0.120	0.990	6.339	1.406	"	0.838	"
Michigan	7.182	0.887	0.813	3.920	1.063	"	0.802	"

Figure 1



REFERENCES

Eichhorn, W., 1969, "Eine Verallgemeinerung des Begriffs der Homogenen Produktionsfunktion," Unternehmensforschung, 13, 1969.

— 1970, Theorie der Homogenen Produktionsfunktion. Lecture Notes in Operations Research and Mathematical Systems, Vol. 22, Springer-Verlag, Berlin-Heidelberg-New York.

Färe, R., 1973, "On Scaling Laws for Production Functions," Zeitschrift für Operations Research, 17, 1973.

— 1975, "A Note on Ray-Homogeneous and Ray-Homothetic Production Functions," The Swedish Journal of Economics, 77, 1975.

Goldman, S.M. and Shephard, R.W., 1972, "On Eichhorn's Generalization of Homogeneous Production Functions," Zeitschrift für Operations Research, 16, 1972.

Hanoch, G. and Rothschild, M., 1972, " Testing the Assumptions of Production Theory: A Non-parametric Approach," Journal of Political Economy, 80, 1972.

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

— 1970, Theory of Cost and Production Functions, Princeton University Press, Princeton.

— 1974, "Semi-Homogeneous Production Functions and Scaling of Production," Lecture Notes in Economics and Mathematical Systems, Vol. 99, entitled: Production Theory, Springer-Verlag, Berlin-Heidelberg-New York.

Zellner, A. and Revankar, N.S., 1969, "Generalized Production Functions," The Review of Economic Studies, 36, 1969.

Zellner, A., Kmenta, J. and Drèze, J., 1966, "Specification and Estimation of Cobb-Douglas Production Function Models," Econometrica, 34, 1966.

Publications in English
1978

The Importance of Technology and the Permanence óf Structure in Industrial Growth. (Eds G Eliasson, B Carlsson, I Nadiri). Proceedings of a Symposium at IUI, Stockholm, July 18–19, 1977. 237 pp.

A Micro-to-Macro Model of the Swedish Economy (ed G Eliasson). Papers from the Symposium on Micro Simulation Methods, in Stockholm Sept. 19–22, 1977. 240 pp.

On Monopoly Welfare Gains, Scale Efficiency and the Costs of Decentralization. Lennart Hjalmarsson, Booklet No. 88. 19 pp.

On the Effects of Different Patterns of Public Consumption Expenditures. Lars Dahlberg and Ulf Jakobsson, Booklet No. 87. 11 pp.

Approaches to the Theory of Capital Cost: An Extension. Jan Södersten. Booklet No. 85. 7 pp.

Determinants of Housing Demand – Analysis of Census Data for the County of Stockholm, 1970. Gunnar Du Rietz. Booklet No. 82. 14 pp.

Economies of Scale and Technological Change: An International Comparison of Blast Furnace Technology. Bo Carlsson. Booklet No. 81. 23 pp.

Publications in Swedish*
1978

Teknisk utveckling och produktivitet i energiomvandlingssektorn (Technical change and productivity in the energy conversion sector). Anders Grufman. 186 pp.

Industrins tillväxt och långsiktiga finansiering (Growth and Long-term Financing of Industry). Lars Wohlin and Bo Lindörn. Booklet No. 86. 62 pp.

Skattepolitisk resursstyrning och inkomstutjämning. En analys av företagsbeskattning och indirekt beskattning (Income Distribution, Resource Allocation and Tax Policy. An Analysis of Business Taxation and Indirect Taxation). Göran Normann and Jan Södersten. 197 pp.

Substitutionsmöjligheter mellan energi och andra produktionsfaktorer (Possibilities of Substitution between Energy and Other Factors of Production). Lennart Hjalmarsson. Booklet No. 84. 68 pp.

Relativprisutvecklingen på energi och dess betydelse för energiåtgång, branschstruktur och teknologival. En internationell jämförelse (Relative Energy Prices and Their Impact on Energy Comsumption, Industrial Structure and Choice of Technology: An International Comparison). Bo Carlsson. Booklet No. 83. 87 pp.

**Almost all studies in Swedish are published with a short summary in English. A list of all books published with titles in English can be obtained from the Institute on request.*

All the publications can be ordered from Industriens Utredningsinstitut. Grevgatan 34, 5 tr, S-114 53 Stockholm, Sweden