Arbeidsnotater

STATISTISK SENTRALBYRÅ

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 68/24 Oslo, 27 November 1968

A Research Proposal: Production and Behaviour Relations in Norwegian Manufacturing Industries. An Exploratory Cross-Section Time-Series Study.

by

Vidar Ringstad

1. Introduction.

Econometric analyses of production and behaviour relations based on combined cross-section time-services data are very rare indeed. Having looked through what I believe are the more important econometric studies of production carried out the last ten years I found only two studies discussing the theoretical problems of combining cross-section and time-series data for production units.[1] And I found in fact only one study of some size where such data are applied.[2] Thus even if my reading may have been superficial and therefore the probability of having overlooked some important work in this field is rather high it is quite clear that this kind of studies has an ignorable fraction of the vast amount of econometric studies of production. Even the more recent survey[3] of empirical studies of production with its 345 studies in the list of references has only a few lines about combining cross section and time series data and itrefers particularly to the theoretical article of Mundlack.[4]

Now, much has happened inthe field of estimation of production and behaviour relations since this survey article was written, but obviously not in the particular field we are going to consider in this study. One is still applying either time-series or cross-section data. Thus with the outstanding exception of Krishna's study Walter's survey makes a good status of the state of arts as concerns studies based on combined cross-section time-series data.

There is one obvious reason why such studies are rare: The empirical basis necessary for carrying out such studies is rarely available. Thus if you don't have data for it, it isn't very interesting even to discuss theoretically the econometrics of such studies. So even if a combination of the across dimension

1) Y. Mundlak: Estimation of Production and Behaviour Functions From a Combination of Cross-Section and Time-Series Data. In C. F. Christ and others: Measurement in Economics, Stanford 1963. and M. Nerlove: Estimation and Identification of Cobb Douglas Production Functions, North Holland Publishing Co., Amsterdam 1965, Chapter VII and Appendix to this chapter.

2) This is a recent Ph. D. study from the University of Chicago by K.L. Krishna: Production Relations in Manufacturing Plants, An Exploratory Study. Chicago 1967. See, however, also Y. Mundlak: Empirical Production Functions Free of Management Bias. Journal of Farm Economics 1961 and I.Hoch: Estimation of Production Function Parameters Combining Time Series and Cross Section Data Econometrica 1962.

3) A.A. Walters: Production and Cost Functions: An Econometric Survey. Econometrica 1963.

4) See footnote 1) above.

and the time dimension in principle is superior to either of the two [5], the former has not caught much interest among econometricians, and at least not so much interest as it deserves.

The situation in this field is very much the same as the one we have in econometric studies of production in general. One is estimating neoclassical production functions because the data usually available suits such models quite well. And the reason for this may be that these data are very much coloured by a neoclassical way of thinking. If the theory of measurement is neoclassical the measures can hardly be used to anathing else than to investigate neoclassical theories. This does not mean that neoclassical theories are useless, quite the contrary. The point is that these theories have been so dominating the last decades that it has been difficult to investigate alternatives empirically. Non-neoclassical data has been rare and unsystematic and thus the discussion of non-neoclassical models has been limited too. [6]

As I consider the data I'm going to apply as very much neo-classical, I would like to point out, to be sure, that I'm not going to leave the safe world of neo-classical production functions. And to lower the expectations still further: This will be, as the title points out, an exploratory study where I guess most of the fundamental commandments of theoretical statistics are ignored. The purpose of the study is to investigate some well known and some less well known theories of production and behaviour. There will be some fishing in data or data snooping, and therefore we cannot expect the comfor assumptions necessary for applying for instance standard t and F tests without qualifications to

5) This superiority refers to the possibilities of reducing different types of bias present both in time-series studies and cross-section studies. In other words the identification problem is in principle easier to solve in a combined cross-section time-series sample. This will naturally be discussed more detailed in the third chapter of this research proposal where the theoretical basis for the study is sketched.

6) See the list of references of A.A Walters survey article refered to in footnote 3 above. See also the list of references of M. Nerlove. Recent Empirical Studies of the CES and Related Production Functions. in M. Brown: The Theory and Empirical Analysis of Production. Studies in Income and Wealth No 31 New York 1967

be fulfilled [7]. However, the standard statistical tools will be applied as if these assumptions were fulfilled. But if it has some importance for the interpretation of the results I'll admit it when I am sinning against the theoretical statistical commandments provided also that someting essential can be said about the possible effects on the results due to the application of theoretical statistical tools in inappropriate situations.

What econometricians needs mostly (except better theories and better data!) is a more suitable statistical theory of testing sequences of hypoteses. This is possible in principle if the sequence can be formulated inside the frame of variance-covariance analysis. [8] Otherwise the connection between the different hypotheses is usually too vague and rejection or nonrejection of a hypothesis is more or less based on intuition, or general evaluation of the results. This later 'method' is not without value if we have some additional informations; For instance a priori knowledge about the sign or 'reasonable' level of a coefficient, provided that what we believe is 'reasonable' isn't a prepossession established by misspecified models previously applied.

One thing making 'data fishing' particularly necessary is the everpresent problem of discrepancies between the way variables should be measured and the way they really are measured.[9] Given a certain kind of measurement error one can in some cases say something 'qualitative' about its effects on the estimates.[10] But to learn something about their importance when a certain estimation method is applied on a certain model for a certian set of data we

---

7) It is difficult to find any good excuse for this procedure. But having a given set of data with little or no knowledge of the behaviour and performance of the central informations available about the production units, it is difficult to come very far without some extent of experimenting, that is a sequence of trials with current modification of the plans in light of new results. Now one shall not excuse ones own sins by arguing that they are sins generally committed by most econometricians. But I would like to point out that it isn't, from a theoretical statistical point of view, worse to learn by ones own doing on the same set of data how to improve the performance of the models than to learn by others doing on the same set of data (for instance on data from U.S.Manufacturing 1957); one researcher learning of the previous how to improve the results.

8) The possibilities to carry out such sequences of tests will be considered in the theoretical chapter.

9) This does not imply that we know the 'correct' measure of a variable if any such exists, but that we know that there on the basis of the theory of production is a better measure, or better measures than the one applied.

10) For instance when undeflated numbers are used in a case when deflated numbers should be used according to the underlying theory.

have usually to try the method on the model for the set of data under consideration. By applying different models and estimation procedures with different degrees of roboustness towards measurement errors one can try to get an impression of the importance of the errors introduced, or in other words how serious the identification problem due to mismeasurement is in our sample.

But due to the complexity of the peculiarities of such data as those we are going to apply we can hardly get an undisputable ranking of estimation methods according to degree of roboustness. And the performance of one estimation method does not necessarily tell us much about how other methods will behave. And if different methods are applied on different models we may run into a sort of an identification problem. If one model "works", that is, gives "reasonable results" while the other gives "unreasonable results" we don't know if it is due to the models or the estimation procedure.

The discussion above is in general terms and as such it isn't very useful. But it is an attempt to justify the exploratory nature of this study and if we run into such problems as these mentioned above they will be discussed in their proper context.

The discussion above does also indicate quite clearly that even if the (still unproved) statement presented at the beginning of this introduction is true, that combined cross-section time-series data is superior to either of its components it is quite obvious that estimating production and behaviour relations from that kind of data we are going to apply is an uncomfortable task. As I am going to do it in spite of the serious limitations both as concerns the models applied and the data, and thus the imperfect knowledge of what is the proper way of estimating the coefficients, it is for three reasons: a) We'll gain some more experience about the performance of combined cross-section time-series data in econometric production function studies. b) I believe, seriously speaking, that we can also learn something about the structure of Norwegian manufacturing establishments which is the empirical basis of this study and c) learn something about the peculiarities of the data, try to identify them, consider possibilities of eliminating them and point out possible other informations than those now available that could be more useful in econometric research of this kind.

The present study is so to speak an extension of a recent study I have contributed to.[11] The latter is based on Census data for Norwegian manufacturing in 1963, and thus it is a pure cross-section study. In the

11) Z. Griliches and V. Ringstad: Economies of Scale and the Form of the Production Function. Forthcoming. This study will be denoted "the Census study" in the following.

present study the time-dimension is introduced, but at the cost of a reduced across dimension, that is fewer units as compared to the Census study [12].

As I would like to have the latter study as a basis for comparison of the results obtained for the present one I have firstly tried to define the variables in the same way, to the largest extent possible, and secondly I'll also investigate the behaviour of some of the models applied in the Census study. But, naturally, I'll also include other models that are possible to analyse due to the time-dimension in our data.

In this draft of the frame of the planned study I concentrate on the empirical basis with a presentation of the informations available, variable definitions and certain empirical problems. This is the contents of the following section. In the next section I discuss the estimation of simple models when having combined time-series cross-section data, I also have some suggestions how to trace the effects of such things as embodied technical change, adjustemnt costs and transitory variations in demand, and some considerations about the form of the production function, particularly the problem of substitution, as I would like to consider the possibilities to analyse some questions discussed in the Census-study.

The theoretical discussion is naturally not complete. It just indicates where I intend to begin. When the ideas in this chapter are tried it will probably lead to reformulations of the models and also to investigations of the performance of other methods of estimation. I have no line of action from beginning to end ready at this stage. But if the simple estimation methods applied on simple models does not give fairly reasonable results it is doubtful if more refined methods applied on the same models work, even if they in principle are more robust against certain types of mis-measurement or mis-specification. That they should work on more complex models is in such a case still more improbable. We are then very probably in a situation where the idenfification problems cannot be solved or reduced whatever we do. But if the first trials work we may have a first approximation from which we start the search for better models and more appropriate estimation procedures. That is, by means of what we have learnt about the behaviour of our data and the performance of our models and estimation methods, we look for a a better approximation.

12) The Census study covered 5351 production units while the present one covers about 900. See next section.

2. *The Empirical Basis of the Study.*

a) The data available:

The units of the present study are the establishments of Norwegian firms in Mining and Manufacturing which in 1963 had at least 100 persons employed on the average.[1] This makes about 600 firms with somewhat more than 1300 establishments. This selection was made on the basis of the 1963-census of Mining and Manufacturing. By means of the Annual Industrial Production Statistics, data for the establishments selected were also obtained for the years 1959-62 and 1964-67. This implies that a) establishments of large firms (according to our definition) dissolved between 59-62 are not registered at all. b) establishments of large firms founded after 1963 are neither registered.

We have also in this raw set of data a number of incomplete time-series for the period under consideration. For some purposes it is particularly interesting to study a mix of complete and incomplete timeseries, for instance to investigate the "birth and death" aspect of the structure in a country's manufacturing industries.[2] But for our purpose we would like to have complete timeseries only. And therefore those establishments that are not registered in one or more of the years 1959-67 were excluded.[3] In addition there are among the establishments selected a number of auxiliary units and investment establishments[4] which have registered employment but no production. They are naturally also excluded, due to the purpose of our study. We have then about 900 complete time series of production units left for the period 1959-1967.

These 900 units will be divided into a number of industries, how many and in what way is not yet decided. It is, however, quite clear that our industry definitions have to be very rough as the total number of units is quite low. I will probably construct 5-10 industries with the two-digit ISIC-code as the basic unit of classification. On this point there is a possibility of testing, that is specifying a fairly large number of industries and try to group the industries according to the results of a sucessive testing. This possibility will be seriously considered.

In addition to general characteristics such as identification number, industry-group, location, type of ownership, we get for each establishment for each

---

1) Number of persons employed in the firm is defined as number of production workers + number of employees + number of owners and family numbers working dayly in the establishments of the firm.

2) A study of this and related problems in Norwegian manufacturing industries is: F. Wedervang. *Development of a Population of Industrial Firms*. Bergen 1965.

3) This implies that we probably also exclude some establishments that have been in operation the whole period if they have changed identification-number for one reason or another. It is, however, difficult to guard against such things and it would imply a heavy amount of work to check such changes in identification.

4) Investment establishments are new establishments not fully "established" in the sense that they have not yet started production in the year for which we get the informations. Such establishments are however automatically excluded when excluding incomplete time series, except for very extreme cases.

year the following informations that in one way or another will be used directly when constructing regression variables

	X_1	Production on own account
1)	X_2	Repairs
	X_3	Contract work
	M_1	Raw materials
	M_2	Packing
2)	M_3	Fuel
	M_4	Ancillary materials [5)]
	M_5	Contract work [6)]
	n_1	Number of production workers
3)	n_2	Number of employees
	n_3	Number of owners and family members
4)	h	Number of hours, production workers
	W_1	Wages, production workers
5)	W_2	Wages, employees
	W_3	Wages, home-workers
6)	U_1	Duties
	U_2	Subsidies [7)]
7)	I_1	Investments, purchased capital goods
	I_2	Investments, repairs and maintainance

---

5) For the year 1959-60 we have informations about $M_3 + M_4$ and for 1965-67 M_4 is included in M_1. Thus only for the years 1961-64 do we have separate informations about $M_3 - M_4$.
6) Except for 1959-1960 we also have informations about traded goods. There is an argument of including traded goods bought and traded goods sold among our input and output components listed. But as we do not have these informations for all years they are ignored.
7) Informations about duties and subsidies are not available in 1959 and 1960. (See below)

H_1 Inventories, raw materials
8) H_2 Inventories, goods in process
H_3 Inventories, finished goods

In addition to these informations we have for the years 1959 and 1963 also informations about:

K_1 Value of buildings
9) K_2 Value of machinery [8)]

By means of the informations above we will try to construct the variables needed for the present study.

It should not be necessary to stress the limitations of the informations available for the production units under consideration. It is, however, not possible for us to cure all defects in this context. But we consider it highly desirable to take care of the variations in the price level of output and input of raw materials. This is impossible to do for each establishment but we may have additional informations about price-variations that may be of some value even if they are not "establishment specific".

The best informations we can get about variations in the price level we have in the more disaggregated national account system.[9)] There we have informations about gross production value and value of raw-materials (that is, input from other sectors, domestic and foreign) in current and constant prices. Using these numbers to construct a "price-index" for gross production and one for raw material consumption we can deflate the two components of value added separately and we can, thus, for each establishment construct a price index for value added.[10) 11)] The numbers from the national account system we have used are naturally computed by means of informations from the same source of data as is the basis of the present study, namely the Annual Industrial Production Statistics. The main difference is that while our sample consists of establishments of large firms only, the national account sectors cover all establishments of the Annual Industrial Production Statistics. The question "how good" these indices are is therefore very much depending on if the product mix of the establishments of our study is much different from the product mix of those of the same national account sector not included in our sample. If there have been only small changes in relative prices of output and input of raw materials and semiproducts this problem is naturally

8) For 1959 we have also informations about "other property".
9) For Mining and Manufacturing there are about 35 sectors.
10) See below.
11) The base year in the national account system is now 1961. For the years 1959-61 the base was 1955. By means of the price indices in 1961, with base in 1955 we can also compute price-indices in 59 and 60 with base in 1961.

not so important.

One may of course as many researchers do, question the general validity of indices in this context. That is trying to squeeze multiproduct multi-input production into a one product - "few"-input frame. This is a rather serious and very important problem which has been superficially treated in econometric litterature. To deal with it in a satisfactory way we probably need a multi-product production function approach. Discussion of this approach falls partly outside the scope of this study. But what could be of some interest is to see what can be done with it in light of the detailed informations about output of different goods and consumption of different kinds of raw material we have in the Annual Industrial Production Statistics. These informations are rather difficult to get hold of as the only informations kept over any longer period is the sum-values of production and input presented above. It seems, however, to be possible to get such informations for the period 1963-67. If this is true and the informations about input and output of all kind of goods are considered to be of fairly good quality it is probably worth the effort to make an excursion into the problems of input and output indices, or multiproduct production functions, by means of this empirical base.

This is something to be more seriously considered at a later stage of the study. Particulartly one has to consider if it should be studied separately or if one should try to analyse it for the about 900 units of this study as we for these also have other informations that makes a more complete multiproduct production function study possible.

Also for capital and investments we need deflators. In this case it is even still more difficult than in the case of gross production and raw material consumption to get deflators that take care of the individualities of the units of our study. The price-indices available refer to total Mining and Manufacturing. Thus we cannot take care of differences in price-movements of capital in different industries. There are indices for different types of capital, but as we have no possibilities to take care of the composition of capital for other years than 1959 and 1963 we must in fact use one priceindex for the two kinds of capital we have, namely buildings and machinery. What we can hope for is that this index take care of the general movement of the price of capital over time. Differences across units or even across industries cannot be taken care of.

In addition to the two main components of capital we have also inventories that is often introduced into the capital measure. But this does not seem to be convenient to do in this case, it is probably more relevant if we could construct a service of capital measure, that is, a measure with different weights of the different components in opposition to the total value of capital which is an unweighted sum and which is the measure to be applied by us. But if we at a later stage would like to include one or more of the components of inventories into the capital measure, and particularly as they will be used in other contexts in this study they should be deflated. And it seems then to be convenient to use the price index of raw-materials to deflate the inventories of raw materials and the price index of gross production to deflate inventories of goods in process and finished products.

As we would like to make the results of this study as much as possible comparable to the Census-study we try to follow the definitions of that study as far as possible.

First of all we have the gross output measure in sellers prices

$$(10) \qquad Y' = X_1 + X_2 + X_3 + U_2 - U_1$$

The input of raw materials etc. is measured as

$$(11) \qquad M' = M_1 + M_2 + M_3 + M_4 + W_3$$ [12]

These are both in current prices, and by deflating with the price-indices discussed above we get gross production and raw material consumption in 1961 prices for all years. Denoting these two indices P_Y and P_M respectively we have:

Gross production in 1961 prices

$$(12) \qquad Y = \frac{Y'}{P_y}$$

and raw material consumption in 1961-prices as:

$$(13) \qquad M = \frac{M'}{P_M}$$ [13]

We get the value added in current factorprices as:

$$(14) \qquad V' = Y' - M'$$

and value added in 1961-prices as

$$(15) \qquad V = Y - M.$$

---

12) Note that in addition to deliveries from other production sectors we include in our raw material input measure also wages to home-workers.

13) W_3 is therefore also deflated with this index. But as W_3 is usually rather unimportant this deflating procedure should not introduce too much error in our measure.

Thus we have also implicitly defined a "price-index" for value added in sellers prices; namely:

$$(16) \qquad P_v = \frac{V'}{V} = \frac{Y' - M'}{\frac{Y'}{P_y} - \frac{M'}{P_m}} = \frac{P_y Y - P_m M}{Y - M}$$

Labour input will be measured in the following way:

$$(17) \qquad L = \frac{h(W_1 + W_2)}{W_1} + 2n_3$$

That is: First we compute labour input of employees in production worker hours equivalents. Second we assume that owners and family members work 2000 hours a year on the average. This is approximately the average number of hours worked by production workers in Mining and Manufacturing in 1963.

In the Census-study it was found that on the average this seems to be an over-statement of the work done by owners and family members. We should therefore probably give n_3 a somewhat smaller coefficient in the labour input measure above. An alternative to this is to introduce an additional variable to investigate to what extent we in fact have overstated (or understated) the work done by this type of labour power. As we in no case know what is the "true" coefficient of n_3 in our labour input measure this alternative is probably somewhat better.

As we may be interested in analysing, the effects of the two main components of labour input separately we may simply use

$$(18) \qquad h \text{ and } \frac{h W_2}{W_1}$$

separately. As these do not take care of the third component of labour input we could instead use:

$$(19) \qquad h \text{ and } \frac{h W_2}{W} + 2n_3 \quad \text{or} \quad h \text{ and } n_2 + n_3$$ [14)]

14) These variables were tried in the census study. We could naturally have three separate variables for labour input, one for each component. But as $n_3 = 0$ for most establishments this is rather inconvenient as we mostly are going to operate in logs of the variables.

The price of labour input variable we are going to apply is simply average wages per hour of production workers, that is.

$$W = \frac{W}{h} \qquad (20)$$

It does not seem possible to construct two separate measures for the price of labour input; one for production workers and one for employees. In fact as we measure labour input in the way we have done, average wages of production workers is also the proper "price" of employees too.

The input of capital will simply be measured as.

$$K = K_1 + K_2 \qquad (21)$$

In principle we have possibilities to construct a more refined measure of capital input, but it does seem worth while as we, when trying to make the capital measure more refined, very well may introduce more errors in the measure. One thing worth serious consideration is if we should add inventories to the measure above or not. It is a bit premature to decide on this now. The main problem is that there probably is a significant transitory element in inventories, reflecting transitory variation in demand and probably also transitory variation in supply of raw materials to some industries. Thus the informations available about inventories and the possible applications of these informations should be discussed in a somewhat wider frame than only in connection with the measure of capital input. Below we discuss more detailed the measurement problems we have of capital in the present study.

Those variables presented above are the main variables of the present study. Some more will be applied, and they are introduced and explained in the context into which they enter.

b) Estimating missing observations for capital.

We have two informations about capital: value of machinery and value of buildings. We could make a long list of defects of these capital data when they are applied as measure of the productive performance of the capital input. But it seems to be of no use to discuss such defects we have no possibility to cure. Those defects we can do something to cure, we should, however, discuss more explicitly. And there are three problems connected to our capital data that we are going to discuss in this section.

a) For some units there are either in 1957 or in 1963 missing information either on value of machinery or value of buildings.

b) For some units (but for a much smaller number) both informations are missing.

c) We have informations about capital only for 1957 and 1963. We have, however, informations about investments for all years covered by our study. How can we compute capital values for the other years by means of these informations?

About a) we shall not say much. There are of course many ways of estimating a missing component of a variable. One possibility is simply to compute the average ratio between value of machinery and value of capital for those establishments having both informations, assuming that there is a "normal" ratio between these two components of capital for each industry. Denoting this ratio $\tilde{\eta}=\frac{1}{m}\Sigma\frac{K_1}{K_2}$ we get for unit i which has a missing information of machinery that $\tilde{K}_{1i}=K_{2i}\tilde{\eta}$, and correspondingly for a unit j with a missing information of buildings; $\tilde{K}_{2i} = K_{1i}/\tilde{\eta}$.

An alternative to this method is to exclude units with one of the components missing. Then units with both informations missing will be excluded also, and then problem b) above disappears. For these units we would, however, like to do something else. That is, to consider a method of estimating the missing value of total capital.

Our point of departure is the value added Cobb-Douglas relation and the least square estimation method. We write this function in logs as

$$y = a + \alpha x + \beta z + u \qquad (22)$$

where y = lnV, x = lnL, z = lnK and u is the usual error term. In our case some of the z's are missing,[15] and there are a number of different ways the missing values could be estimated.[16] In our case only one will be considered; that is estimation of the missing z-values by means of the least square method applied on (22).

Having n sets of observations of which n_1 are complete and n_2 incomplete (that is z = 0) and can write the sums of squares function to be minimized with respect to $z_{n_1+1}, \ldots z_n$ and α and β as.

---

15) In fact K=0 and this implies that z = $-\infty$. But imposing $\tilde{K}$ = 1 whenever we observe K=0 we have correspondingly a missing capital value whenever z=0.

16) See R.M. Elashoff and A. Afifi: *Missing Values in Multivariate Statistics, I Review of the Literature, II Point Estimation in Simple Linear Regression*, Journal of the American Statistical Association 1966 and 1967.

(23) $$U^2 = \Sigma^{n_1}(y_1-\ddot{y}-\alpha(x_1-\ddot{x})-\beta(z_1-\ddot{z}))^2+\Sigma^{n_2}(y_2-\ddot{y}-\alpha(x_2-\ddot{x})-\beta(z_2-\ddot{z}))^2$$

where the first sum of squares refers to the complete sets of observations and the second to the incomplete sets of observations.

$$\ddot{y} = \frac{1}{n}\sum^{n} y_i,\quad \ddot{x} = \frac{1}{n}\sum^{n} x_i \text{ and } \ddot{z} = \frac{1}{n}\sum^{n} z_i.$$

First we get the n_2 first order conditions of min U^2:

(24) $$\frac{\partial u}{\partial z_j} = \frac{2\beta}{n}\sum^{n_1}(y_1-\ddot{y}-\alpha(x_1-\ddot{x})-\beta(z_1-\ddot{z})) + \frac{2\beta}{n}\sum^{n_2}(y_2-\ddot{y}-\alpha(x_2-\ddot{x})-\beta(z_2-\ddot{z}))$$

$$-2\beta((y_{2j}-\ddot{y}-\alpha(x_{2j}-\ddot{x})-\beta(z_{2j}-\ddot{z}) = 0 \qquad (j = n_1+1, \ldots, n)$$

As the sum of the two first terms of (24) is zero, due to the properties of the least squeare method we get the n_2 conditions.

(25) $$(y_{2j}-\ddot{y})-\alpha(x_{2j}-\ddot{x})-\beta(z_{2j}-\ddot{z}) = 0 \qquad (j = n_1+1, \ldots, n)$$

That is, all units with incomplete sets of data get a value of z which gives no residual for this unit. This is a rather unfortunate property of our method as we in this way ignore the presence of an error term in the production relation for these units. Or rather, we let our estimates on the capital value "absorb" the error term of the units under consideration. But on the other hand as we do not know a priori the value of the error term for each unit with missing capital values our z's are probably the best we can get. But naturally the "goodness" of these estimates is very much depending on how "good" the single-equation least square method is in this connection. If it gives unbiased estimates on α and β it will also give unbiased estimates on the missing capital value.

Thus, our estimates of z_{2j}, $j = n_1+1, \ldots, n$ are such that the second term of (23) becomes zero. To find the least square estimates of α and $\hat{\beta}$ we insert $\ddot{z}$ in the first term of (23) and minimize this term with respect to α and β.

We get $\ddot{z}$ in the following way. From (25) we have

(26) $$\sum^{n_2}(y_{2j}-\ddot{y}) - \alpha\sum^{n_2}(x_{2j}-\ddot{x}) - \beta\sum^{n_2}(z_{2j}-\ddot{z}) = 0$$

But due to the property of the least square method this implies:

(27) $$\sum^{n_1}(y_1-\ddot{y}) - \alpha\sum^{n_1}(x_1-\ddot{x}) - \beta\sum^{n_1}(z_1-\ddot{z}) = 0$$

Thus we get:[17]

$$\bar{z} = -\frac{1}{\beta}\frac{1}{n_1}\sum^{n_1}(y_1 - \bar{y}) + \frac{\alpha}{\beta}\frac{1}{n_1}\sum^{n_1}(x_1 - \bar{x}) + \frac{1}{n_1}\sum^{n_1} z_1 \qquad (28)$$

Inserting (28) into (23) (as the second term of U disappears) we get in fact:

$$U_1^2 = \sum_{i=1}^{n_1}\left(y_i - \bar{y}_1 - \alpha(x_i - \bar{x}_1) - \beta(z_i - \bar{z}_1)\right)^2 \qquad (29)$$

where

$$\bar{y}_1 = \frac{1}{n_1}\sum^{n_1} y, \quad \bar{x}_1 = \frac{1}{n_1}\sum^{n_1} x \text{ and } \bar{z}_1 = \frac{1}{n_1}\sum z$$

Thus, by ignoring the incomplete sets of data we can get the least square estimates on α and β, that is by minimizing (29) with respect to α and β. We can then go back to (25) and estimate the missing z-values by:

$$\hat{z}_{2j} = \frac{1}{\hat{\beta}}(y_{2j} - \bar{y}_1) - \frac{\hat{\alpha}}{\hat{\beta}}(x_{2j} - \bar{x}_1) + \bar{z}_1 \qquad (30)$$

Of (30) we also see that the intuitive method of inserting the average value of a variable of the complete set of data ($\bar{z}_1$) where it is missing is recommendable only if the values of the other variables for the units with incomplete sets of data is equal or near the average of these variables in the complete sets of data. A method that is somewhat better in our case would be to write the production function as:

$$y - x = a + (\alpha + \beta - 1)x + \beta(z - x) \qquad (31)$$

and estimate the missing log capital/labour ratios (z − x) by the average of this variable for the complete sets of data. The formula (30) is valid also in this case and we see that if we have a linear homogeneous production function this may be a satisfactory method, but the validity of it still depends on if the average productivity of labour (y − x) is much different for those units with incomplete and complete sets of data.

As pointed out above a number of other methods could be applied. But some of them seems to be better suited to other situations than the one above while

17) We note that $\bar{z}$ is the average of the observed + the estimated z's.

other methods seem to be too refined and computationally complicated to be worth applying in our case.

An alternation to a) to be considered is to treat both those units with one component of capital missing together with those with both components missing along the lines sketched above under b).

As concerns point c) our problem is to use the available informations about capital in 1959 and 1963 and the informations about investments to compute capital values for the other years covered by our study. The first sub-problem we have to deal with is then if we should use both categories of investment, repairs and maintainance, and purchase of capital goods. This is something that is difficult to decide at this stage of the study but I am inclined to include the last category only, even if maintainance of the capital stock is very important for the production performance of this factor of production. But this is something to be considered more seriously later. The estimation procedure sketched below is valid, with some minor modifications even if we include repairs.

The second problem arises because the reported capital values of 1959 and 1963 refer to the full fire insurance values at the end of the respective years, while the investments informations is the accumulated flow of purchases of capital goods during the year. The question is if and to what extent the investments carried out during the year is reported as part of the capital value at the end of the year. A recent study gives some indications of a certain lag or sluggishness between reported investments and reported capital. That is, either because some of the investment projects are not completed before next year (or even still later) and the value of the investements carried out during the year is reported while it is not registered as part of the capital stock before it is completed, or there may be a certain sluggishness in adjusting the fire insurance value for new capital goods.

If the value of the incompleted investment projects are not reported as part of the capital stock at the end of the year this is in fact an improvement of the capital measure, but the presence of such incompleted investment projects makes the computations of capital-values by means of reported investments difficult, as we clearly should know how much these incompleted investmentprojects make of the total investments. The sluggishness in adjusting the capital stock for new capital goods makes matters difficult for us in another way, as this implies that the reported capital values of 1959 and 1963 are systematically too low.

---

18) V. Ringstad and Z. Griliches: *"A Method of Analysing Consistency Between Time-series for Capital and Investment"*. Review of Income and Wealth. No 4 Dec. 1968.

This discussion has no other value than to point out certain difficulties when using reported capital and investment data of the kind to be applied in this study . I feel that it is difficult to do anything better than what we probably would have done in any case, namely to compute the capital value of a year as the depreciated capital of the previous year pluss the investment during the year. [19)]

To make the capital and investment data comparable we have to deflate them to bring them to the same price basis. This is another weak point in connection with the constructed capital values as the only deflator we have is one price-index for Total Manufacturing for both categories of capital. In this way we cannot take into consideration different price-movements for different industries due to differences in the composition of capital and investments. There exists price indices for both categories of capital but what we are missing are informations about the composition of investments on different types of capital goods.

Having deflated the numbers to a common price-base, and denoting K the capital stock, I purchase of capital goods and δ the depreciation ratio we get the capital values of the different years as:

$$(32)\qquad \begin{cases} K_{60} = (1-\delta)K_{59} + I_{60} \\ K_{61} = (1-\delta)^2K_{59} + (1-\delta)I_{60} + I_{61} \\ K_{62} = (1-\delta)^3K_{59} + (1-\delta)^2I_{60} + (1-\delta)I_{61} + I_{62} \\ K_{63} = (1-\delta)^4K_{59} + (1-\delta)^3I_{60} + (1-\delta)^2I_{61} + (1-\delta)I_{62} + I_{63} \end{cases}$$

The final equation will be used to estimate the depreciation ratio δ which is assumed to be the same for all units of an industry for all years.

$$(33)\qquad K_{63} - I_{63} = (1-\delta)^4K_{59} + (1-\delta)^3I_{60} + (1-\delta)^2I_{61} + (1-\delta)I_{62}$$

By means of a nonlinear estimation method we estimate the parameters of the relation:

$$(34)\qquad K_{63} - I_{63} = a_0 + a_1K_{59} + a_2I_{60} + a_3I_{61} + a_4I_{62}$$

---

19) There are other objections to this method than those mentioned above. For instance do we assume that the production performance of a capital good is reduced with a constant fraction each year, independent of the age of the capital good. It seems to me, however, to be difficult to apply an alternativ approach that is obviously better than the one proposed.

given the following four constaints:

$$(35)\quad \begin{cases} a_0 = 0 \\ a_1 = a_2 a_4 \\ a_1 = a_3^2 \\ a_2 = a_3 a_4 \end{cases}$$

The relation (34) will probably behave better if we ignore the constaints in (35), that is including a "year effect", and this will also be done, but only to see if the results becomes substantially different from those obtained when (35) is taken into consideration. The reason why we impose the same depreciation ratio for these years is because in will be used when computing the capital values of the remaining years as:

$$(36)\quad \begin{cases} K_{64} = (\widehat{1-\delta})K_{63} + I_{64} \\ K_{65} = (\widehat{1-\delta})^2 K_{63} + (\widehat{1-\delta})I_{64} + I_{65} \\ K_{66} = (\widehat{1-\delta})^3 K_{63} + (\widehat{1-\delta})^2 I_{64} + (\widehat{1-\delta})I_{65} + I_{66} \\ K_{67} = (\widehat{1-\delta})^4 K_{63} + (\widehat{1-\delta})^3 I_{64} + (\widehat{1-\delta})^2 I_{65} + (\widehat{1-\delta})^3 I_{66} + I_{67} \end{cases}$$

where all numbers refer to a common price-base.

If estimating capital values between 1959 and 1963 is risky, estimating capital values for years after 1963 by (36) is rather hazardous. But if we shall use a capital measure in our study at all there seems to be no obviously better alternative to the method above. In any case it should be tried. If it does not work we have to manage without a capital measure, except for those years for which we have a direct measure of this variable.

3. The Theoretical Frame of the Study

a) General Analysis of Cross Section-Time Series Data.

When trying to estimate the parameters of a relation, we can never expect to obtain completely unbiased estimates, whatever method is applied. We'll always face fundamental identification problems when applying non-experimental data with little or no choice as concerns definitions of variable-measures and no control of the accuracy of the informations available - and usually also having limited knowledge of appropriate models. And naturally most of the discussion in econometrics is about specific types of problems present when trying to obtain reliable informations about the parameters involved.

Depending on the particular situation one may use reduced form estimation, two stage least square estimation or instrumental variables estimation to mention a few of the more wellknown methods. Or one may simply estimate the parameters of a relation by means of the ordinary least square method, - using so to speak the right side variables as instruments for themselves. Apart from some naive "regressors" this is usually not done because one really belives that this is <u>the</u> proper way of estimating the variables, but because it is considered to be the <u>more</u> proper method under the given circumstances.

This point of view should also be adopted for the other methods mentioned too: We cannot in general expect to have solved the identification problems completely. To narrow the scope of the discussion, let us consider the methods mentioned above, except the reduced form estimation as that one implies generally estimation of the parameters of more than one relation.

For all three methods under discussion we can partion the estimates obtained into a "systematic part" which is the parameters themselves an a "random" part;

(1) $\hat{a} = a + Mu$

where $\hat{a}$ is a vector of estimates, a the corresponding vector of parameters and Mu is a vector of "random" elements.[1]

The identification of the a-vector depends therefore on the behaviour of Mu. Complete identification implies that the expected value of this term disappears, at least asymptoticly. Now, if we have chosen the more proper method in the sense that the impact of Mu is minimized the question is; can we do still better? Can we reduce the impact of the "random" term somewhat?. This is obviously an empirical question that has to be evaluated on a theoretical basis; we have to use our knowledge of economic theory on what we know about the data.

If we for instance, by means of theoretical considerations and our knowledge of the behaviour of our data, can manage to classify our sample in such a way that the expected value of u is approximately the same for units within each class but obviously different between classes we may reduce the effect of Mu on our estimates by taking care of the differences in expected values of u between classes in the regression computations. Clearly the usefulness of such a procedure depends very much on the importance on our estimates of the "random" term in (1). The procedure consumes degrees of freedom and if the systematic impact of Mu on $\hat{a}$ is low we may very well loose in efficiency due to a lower number of degrees of freedom what we gain in unbiasedness. On the other hand we may also manage both to improve the efficiency and reduce the bias by this method. It depends on the importance of Mu as well as on our ability to classify the units properly.

The discussion above is intended to cover the basic idea of what is usually denoted analysis of covariance. How can this method be applied in our case? We'll discuss it by means of a particular type of production function and a particular type of behaviour relation. But the method is also

---

1) The "contents" of M is naturally different for the different methods. For direct least square estimation we have $M = (X'X)^{-1}X'$ where x is the nxk matrix of the rightside variables of the relation we consider where n is the number of observations and k is the number of parameters. For the two stage least square method we have correspondingly $M = (\hat{X}'\hat{X})^{-1}\hat{X}'$ where $\hat{X}$ is the nxk matrix of estimates X-values from the "first stage" computations. And for the instrumental variable method we have $M=(X'P)^{-1}X'$ where P is the nxk matrix of the instrument-variables. We can naturally use mixed methods, i.e. for instance apply two-stage least-square or instrumental variable technique for only some of the right-side variables of the relation we consider. The discussion below is valid also for such cases.

valid for other types of relations.

We assume that the value added Cobb Douglas function is an appropriate description of the production structure.

Thus we have:

(2) $$V_{it} = AL_{it}^{\alpha} K_{it}^{\beta} e^{u_{it}}$$

where V is value added, L is labour input, K is capital input and u is an error term. A, α and β are parameters to be estimated. The subscripts i and t refer to establishment and year respectively.

In our case there is an apparent way of grouping the units, namely by a cross-classification along the establishment and time dimension respectively. Correspondingly we can decompose the error term:

(3) $$u_{it} = v_i + v_t + v_{it}$$

where v_i is the effect of left out characteristics that in general are specific for each establishment [2)] and correspondingly v_t is the effect of left out characteristics of each year covered by the study. v_{it} is that part of the error term that we hope will not have any systematic variation either along the across-dimension or along the time dimension.

If we had indices of the left out characteristics along the across dimension and time dimension respectively we could have estimated the parameters of the relation:

(4) $$V_{it} = AD_i^{\gamma_1} D_t^{\gamma_2} L_{it}^{\alpha} K_{it}^{\beta} e^{v_{it}}$$

provided that they enter in the form assumed.

As we don't have such indices we have to apply a less restrictive but more degrees of freedom-consuming method, namely the previously discussed covariance, or dummy -variable method. We take care of the systematic differences in u_{it} along the across and time dimensions by the following relation:

(5) $$V_{it} = Ae^{\sum_{j}^{n} a_j d_j} e^{\sum_{\tau}^{f} b_{\tau} d_{\tau}} L_{it}^{\alpha} K_{it}^{\beta} e^{v_{it}}$$

2) Mundlak interpret it as management and in fact does also try to estimate it. But it is not difficult to think of other factors that can be establishment - specific. Y. Mundlak: Empirical Production Function Free of Management Bias. Journal of Farm Economics 1961.

where $d_j=1$ when $i=j$ and zero otherwise and $d_\tau=1$ when $t=\tau$ and zero otherwise. Or written in logs we get (5) as:

$$(6) \qquad \ln V_{it} = \ln A + \sum_j^n a_j d_j + \sum_\tau^T b_\tau d_\tau + \alpha \ln L_{it} + \beta \ln K_{it} + v_{it}$$

As we have pointed out previously in this section this procedure can be applied, whatever method we apply to estimate α and β, either it is ordinary least squans on (6), the instrumental variable method, the two stage least square method or a mix of these methods.

I do not consider it worth while to give the coefficients of the dummy-variables a particular interpretation. Mundlak [3] argues that the establishemnt dummies express differences in management. This may be true in his case, in the sense that variations in management ability "dominates" these coefficients. In our case, at least it is quite clear that there are also other factors involved such as mismeasurement of the variables even if much of it presumably is left in the v_{it}-term.

The interpretation of the time-dummies is probably somewhat easier as it should be "dominated" by technological change. But, of course, it will also to some extent reflect business cycles variations as they are not taken care of explicitly by us.[4]

As a simple example of a behaviour relation let us consider the 1. order condition of profit maximum with respect to labour, when the form of the production function is slightly more general than (2), namely of the CES-type.

$$(7) \qquad \left(\frac{V}{L}\right)_{it} = BW_{it}^{c} \cdot e^{\omega'_{it}}$$

where W_{it} is the real wage-rate for establishment i in year t.

As for the production function the error term ω'_{it} of the behaviour relation may have components that are establishment specific and time specific. For instance it can reflect differences in the behaviour pattern of different establishments and differences in the degree of adjustment to the appropriate real wage rate, different for different years. And also for the behaviour relation we may think of mismeasurement of the variables entering the relation as components of the error term contributing to a certain systematic variation of it along the two dimensions of our study. Thus also for the behaviour relation we decompose the error term into an establishment specific, a year specific and an unsystematic component.

3) See footnote 2) above.

4) See, however, later when we discuss "transitory" variation in demand.

(8) $\omega'_{it} = \omega_i + \omega_t + \omega_{it}$

And in the same way as for the production function we introduce dummy-variables to take care of the effects of the two former components. Thus writing (7) in logs we get:

(9) $\ln(\frac{V}{L})_{it} = \ln B + \sum_{j}^{n} \alpha_j d_j + \sum_{\tau}^{T} \beta_\tau d_\tau + c \ln W_{it} + \omega_{it}$

where as previously $d_j=1$ when $i=j$ and zero otherwise and $d_\tau=1$ when $t=\tau$ and zero otherwise.

The production function (6) and the behaviour relation (9) is the point of departure as concerns functional forms. A number of alternatives will be tried if it on the basis of the first results is considered to be worth while. The ordinary least square method is the point of departure as concerns estimation method. Other methods mentioned at the beginning of this chapter will also be tried, provided ~~still~~ it is considered worth while. But whatever forms of the relations or methods of estimation we apply it is in general convenient ot investigate the effects of the dummies introduced to see if the estimates on the main coefficients become much different and if the fit is improved or not. It is of particular interest to discuss differences in estimates on the main coefficients obtained by means of different estimation methods, and with and without dummyvariables to see if we can give these differences particular interpretations.[5]

---

5) See the article of Mundlak refered to in footnote 1) of the first chapter of this note.

b) An example of testing of successive hypotheses.

If we are willing to assume that the residuals of the production and behaviour relations are normally and independently distributed with constant variance we can test the significance of our across effects and year effects. Now, one might be interested also in investigating a less degrees of freedom-consuming method of taking care of these effects.

As we are going to construct a limited number of industries by means of the ISIC code we instead of establishment effects can introduce industry-group effects by means of the dummy variable technique. And along the time dimension we can instead of including dummies for years include a trend term $e^{\mu t}$. By denoting the establishment effects as "E effects", the year effects as "T effects", the industry-group effects as "I effects" and the trend-effect as "t effect" we get the following mixed sequence of effects, ordned in increasing degree of restrictiveness.

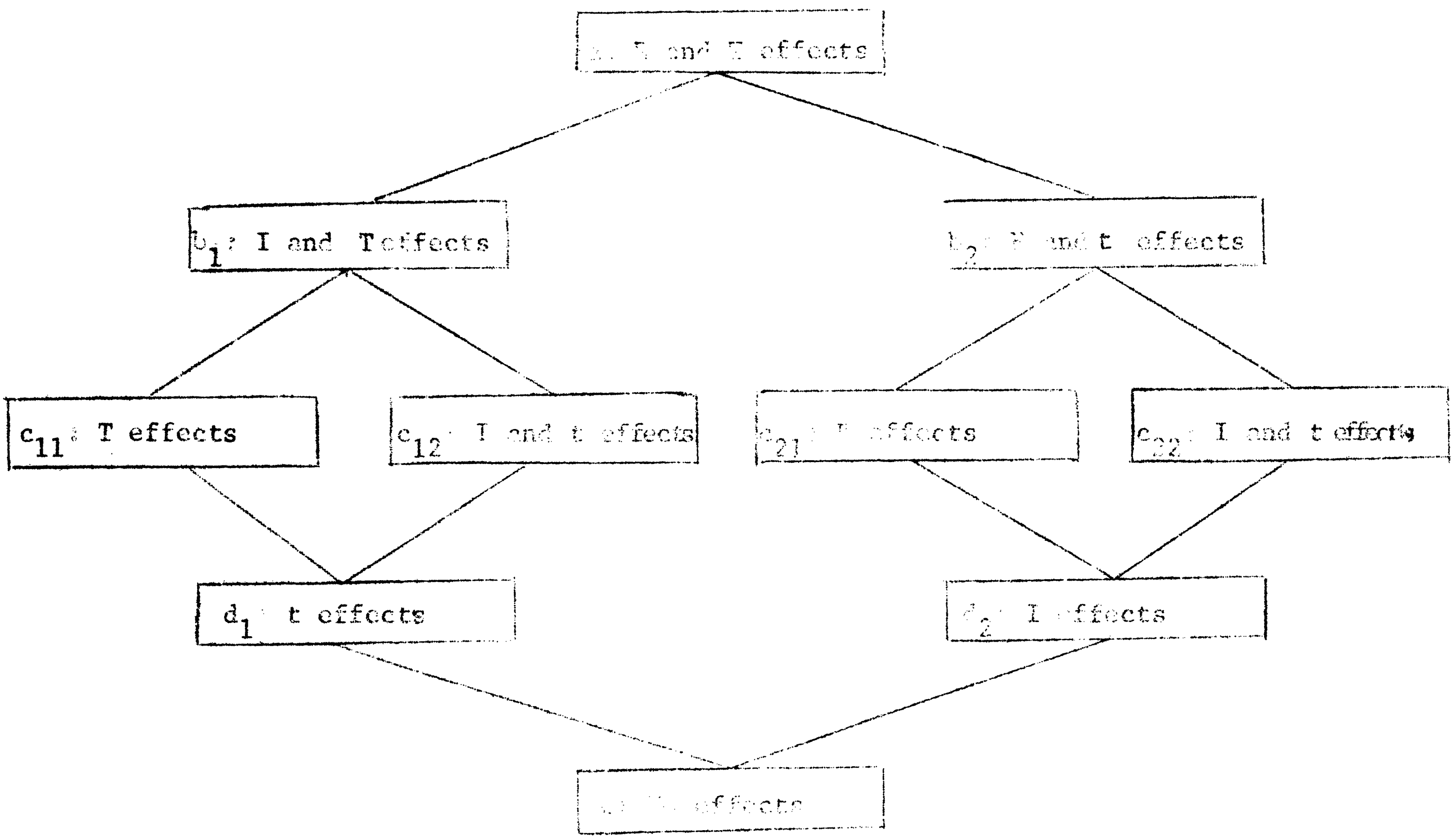

Fig. 1

By means of the results of the mixed sequence of regressions we should get at fairly good impression of the importance and behaviour of the two types of effects discussed. Thus we intend to carry out all regressions implied by the figure above. As the two main types of effects discussed at least partly have different causes in the production relation and the behaviour relation, both types of relations will be estimated with the different assumptions about the nature of the effects.

What kind of conclusions can be drawn from the sequences of tests that can be carried out according to the figure above? If both b_1 and b_2 are rejected it implies that we should go along with the [illegible] and T effects. If, say b_1 is rejected but not b_2 but both c_{21} and c_{22} are rejected (b_2 is one-hypothesis) we go along with [illegible] and [illegible]. The potential number of conclusions is as we see rather high, and it may in some cases be difficult to figure out the proper conclusion to be drawn. If for instance both b hypotheses and all four c-hypotheses have not been rejected but d_1 is rejected when c_{11} is the one-hypothesis but not rejected when c_{12} is the one-hypothesis, and d_2 is rejected when c_{22} is the one hypothesis but not rejected when c_{21} is the one hypothesis - what then? Well then the conclusion is that we either shall go along <u>either</u> with dummies for years <u>or</u> dummies for industry-groups and a trend-term. The test-procedure does not solve the problem of choice we have here, and we must make a decision by means of other criterions. Possibly such a result indicates that our less restrictive model, "a", is not quite well specified. That there in addition to a pure across effect and a pure time effect also is an interaction effect, that is the v_{it} and w_{it} terms of the production and behaviour relation respectively are not quite randomly distributed over all units.

The number of arrows in fig. 1 is equal to the potential number of tests, that is 12. If we like to have an overall level of the sequence of tests of ε we must have a level of each test of $\frac{\varepsilon}{12}$. In fact then ε is a upper bound of the true level as the overall level is lower if the tests are idependent, namely $\varepsilon_1 = 1-(1-\frac{\varepsilon}{12})^{12}$. It could be shown that in this case ε_1 is only slightly lower than ε. In our case we consider it convenient to use a level of 0.005 of the individual tests and thus we get an upper bound of the overall level of the sequence of tests as 0.06.

c) "Long run" versus "short run" production functions.

The interpretation of the results of a relation depends very much on what kind of sample that is applied. For cross-section data one prefers to interpret the results as "long run", while time-series data usually leads us to interpret the results as "short run". The question is then: What is the proper interpretation of our results when using combined cross-section time-series data. In general it will be a "general" function in the sense that it is neither short run nor long run. This is a bit unsatisfactory and we would like to know if we can use our data in such a way that we can say something about the short-run properties of the production process and the long run properties, and particularly if there are substanital differences. And if so, what is the interpretation of these differences.

Thus to study the short run/long run aspect of production relations, we must estimate the parameters of our relations by means of the proper kinds of data.

Krishna [6] in his study asserts that when using the combined cross-section time-series data one gets short run production function estimates. This is, however, not generally true even if one, as Krishna does takes out the inter-establishment and year differences by means of dummy-variables.

What I think is the only valid procedure to get "fairly pure" short-run relations in combined cross-section time series data is to use the averages of the variables over establishments for each year as "observations". That is, if we believe that the Cobb Douglas relation is valid, to estimate the parameters of the relation

$$(10) \qquad \overline{\ln V}_t = a_0 + \alpha_1 \overline{\ln L}_t + \alpha_2 \overline{\ln K}_t$$

where the variables refer to the averages over establishments in year t. By this method we have probably eliminated some of the measurement errors present in our variables. On the other hand we have no measure of the business cycle variations. But we'll later discuss some possibilities to cure this defect. And if it seems to be promising we'll also modify relation (10) to take care of this effect too. Having only 9 years we must, however, economize with the degrees of freedom and we cannot "extend" (10) too much without running into degrees of freedom problems. What will be done in this direction depends therefore what will be the outcome of a number of excursions we are

6) See reference in footnote 3) of chapter 1.

going to pursue inside the general cross-section time-series framework.

It is of course not only for production functions it is of interest to look at, and compare short-run and long run elasticities. The same is also valid for behaviour relations. In this case we'll in the first round limit ourselves to consider the ACMS relation. That is we'll estimate the parameters of

(11) $\qquad (\overline{\ln\frac{V}{L}})_t = a+b\overline{\ln W}_t$

By applying time series of averages of variables for industries we have probably reduced possible biases in the estimate on the elasticity of substitution due to errors of measurement such as differences in quality of labour and incompletely deflated production and wage-variables across establishments. On the other hand the behaviour implicitly assumed to be in operation when applying (11) may be rather doubtful. As we also for behaviour relations are going to consider modified versions of, or related version to (11) we may also feel the necessity to specify the short run behaviour relation differently from the specification in (11). But as for the short run production relation the number of degrees of freedom may quite probably become an effective constraint on the possibilities of extensions and/or modifications of relation (11) above.

As concerns the long-run elasticities case we can do something related to what is done in the case of the short-run elasticities above, namely to compute the estimates by means of averages of the variables, but averages this time computed along the other dimension of our data. That is, we compute averages of the variables over the nine years of our study for each establishment. Thus analogous to (10) we now estimate the long-run elasticities of the Cobb Douglas relation by:

(12) $\qquad (\overline{\ln V})_i = \alpha_0+\alpha_1\overline{\ln L}_i+\alpha_2\overline{\ln K}_i$

and the long run elasticity of substitution by

(13) $\qquad (\overline{\ln\frac{V}{L}}) = a+b(\overline{\ln W})_i$

What is said above about modified versions of (10) and (11) and the possibilities of alternative specifications is also valid for (12) and (13) too, but in this case there are no real degrees of freedom problem if we consider it convenient to include other variables in our long run relations above.

Krishna [7] considers also antoher way of estimating the long-run elasticities. That is to estimate the parameters by means of the relative growth of the variables from the first to the last year covered by the study. This method depends heavily on how 'normal' these years are. To my opinion this method cannot be generally superior to the one sketched above (which is also applied by Krishna) as it excludes most of the informations available, namely everything between the first and the last year.

The computations above are clearly not only interesting when considering short-run versus long-run elasticities and 'something between' provided by the general cross-section time series approach. It has also substantial interest when studying different types of specification errors, particularly measurement errors. These things will be discussed in details together with the presentation of the results.

d) Attempts to trace the effects of costs of change, embodied technical change and transitory variation in demand.

i) Costs of change.

Another ignored field of the econometrics of production is the more specific dynamic aspect of the production processes and the behaviour of the production units. The existing theory in this field is difficult to 'adjust' to the present type of data, and it seems to be difficult to come much further by other, alternative approaches too. Thus in total we cannot come very far in this direction in the present study. Therefore this is not the place to present any detailed discussion of such dynamic econometric models. We'll only adopt that part of the theory of dynamic production models which is based on the idea of 'adjustment-costs' or costs of change. [8] The basic idea of this theory that there are specific costs of changing the scale of production. If an establishment wants to hire new workers, or in particular expand the capital stock (or both) resources like organisation and administration etc. of the establishment have to be allocated to this purpose, resources that

---

7) See reference in footnote 2 of chapter 1.

8) See for instance R. Lucas: Adjustment costs and the Theory of Supply. The Journal of Policial Economy. No 4 1967, M. Nerlove: Estimation and Identification of Cobb-Douglas Production Functions. North Holland Publ. Co. 1965. Chapter VII, and C.D. Hodgins: On Estimating the Economics of Large-Scale Production: Some Tests on Data for the Canadian Manufacturing Sector. Ph. D. dissertation Chicago 1968.

otherwise would have been allocated to the purpose of current production. One can also think of costs of contraction, but the problem is not generally symmetrical, that is the costs of change function may be different for expansion and for contraction.

Together with the assumption of "costs of change" one usually also introduce a long run profit function; a long run profit function is maximized instead of a short run one. Thus it is assumed that one is willing, if necessary, to allocate ressources for adjustment of the scale of operation on the expence of the short-run profit to obtain a maximum of the long run or multiperiod (up to the economic horizon) profit function. But certainly adjustment of the scale of operation is generally necessary even to maximize the short run profit function. The assumption of a long run profit function is introduced to determine the adjustment path given the function expressing the costs of such adjustments.

Anyhow, the question of dynamic behaviour is presumably too refined for our ampirical basis to trace and determine with any reasonable degree of accuracy. But we believe that the less demanding problem of costs of change can still be investigated irrespective of what is the "true" technical and behaviour structure. But clearly there will then be strong qualifications to make about the conclusions that can be drawn from such an analysis.

The question that counts in this context, in the sense that it is the only one we have a fair chance to say anything about, is if such costs of change have any importance for the identification of the parameters of our relation. Or put in another way; if the true behaviour is such that our models are fundamentally invalid.

We have not the informations variable that are necessary to carry out a satisfactory analysis on this point. What we can do and what one ought to do in any case is to look at the effect of variables representing the changes of scale of operation when introduced into the production relation, to see if they have any significant effect and also to see if they affect the other coefficients of the relation significantly. If this is so we have to take the effect of these variables into consideration. If not we can conclude that if the idea of "adjustment costs" has any relevance for Norwegian Manufacturing at all we cannot by means of the informations available to us say anything about it.

There are in this case, as I would say in almost all other cases when trying to get a conclusion from results of an applied econometric analysis, some qualifications to make. As pointed out above, this is particularly true in the present case. And I would like to mention them explicitly.

First, even if our variables are indicators only one must consider them as endogeneous in the production function [9]. This relation is now at best a partly reduced form relation of an unspecified behindlying model. And one can hardly think of such a model where the variables under consideration are exogeneous. Thus we have run into the well-known simultaneous equations problem.

Second, the variables we can think of in this context, expressing change in scale of operation is net investment $I-\delta K$ and $N_t - N_{t-1}$. These variables may reflect also other things than costs of change. For the first variable we may get a negative effect even if there are no costs of change of capital if (some of) the reported investment does not come into operation the same year it is reported. On the other hand we may also get a positive effect of this variable if there are neither costs of change nor new capital goods not in operation, if the embodymenthypothesis [10] has some relevance in Norwegian Manufacturing industries. Some of the same arguments are also valid for the variable assumed to take care of costs of change in labour input. But in any case it should be of interest to "take out" the effect of these variables to make the relation more stable. [11]

Third as we have not specified the model behind the "extended" production model we cannot say what is the proper form in which the variables should enter the relation. It might very well be better to "take out" the scale-effect of these variables: That is, instead of using them as they are presented above we use $\frac{I-\delta K}{K}$ and $\frac{N_t - N_{t-1}}{N_{t-1}}$. This should be an open question at this stage, and a matter of moderate experimentation.

What we can do in this important field is rather unsatisfactory. The variables hopefully manage to catch some of the effects of possible misspecifications of our models due to dynamic factors present in the behaviour of the production units of our study. But as pointed out, both significant and not significant coefficients of the variables introduced for this purpose will be a bit difficult to give a satisfactory interpretation.

---

9) When such variables are introduced this relation can hardly be denoted "a production function" any more. This notation is, however, retained for convenience.

10) The embodyment hypothesis will be discussed briefly below.

11) This statement is a bit doubtful if the variables really are endogeneous: that is if they really can tell us anything about "costs of change".

ii) The embodyment hypothesis.

The embodyment hypothesis has been subject to much interest among growth economists in recent years. The basic idea of this hypothesis is roughly speaking that technological change is initiated through investment in new capital goods, and thus that capital goods of recent vintages are more productive than older ones. The attempts made to verify this hypothesis have not been very successful, however. And I don't think we are better off, empirically, in this study to say much about this subject. It is, however, tempting to try a tentative test on this hypothesis by means of the same method as Krishna applies.[12)]

The method is intuitively plausible and is easy to apply, as it implies an additional variable in the production function only, measuring the 'recentness' of capital. That is we compute a variable expressing the ratio of the recent years investment on the total capital value. Like Krishna we select the investment of the previous three years. This is a bit arbitrary, but taking the degrees of freedom available into consideration we can hardly include investment of a longer period, and including less than three years could lead to a poorer measure if there are substantially different degrees of variation of investment over time of the different production units. As we have a constant depreciation ratio, independent of vintage, year and industry we get our measure of recentness of the capital stock as:

$$R = \frac{(1-\delta)^3 I_{t-3} + (1-\delta)^2 I_{t-2} + (1-\delta) I_{t-1}}{K_t} \qquad (14)$$

We have not included current investment in our measurement because of the qualifications made in the previous chapter about this information.

Success or failure of the approach above depends heavily on the across-dimension of our study, as we cannot expect to get any clear evidence for or against the embodyment hypothesis along the timedimension. We therefore compute the average of our variables for each establishment and use these averages in the regression computations. Because of the definition of R above we 'loose' three years. So we compute our variables from t=62 to t=67.

---

12) See footnote 2) in chapter 1.

The presumably more important qualifications should be pointed out explicitly. The embodyment hypothesis has something to do with a proper measure of the capital input variable. If the hypothesis is true, and of some importance we will expect to get a positive and significnat coefficient of R when it is introduced into the production function. Now a significant coefficient of our R variable may also reflect other weaknesses of our capital measure. We have for instance assumed that the productive performance of a capital good decreases with a constant rate. That is, we have assumed the depreciation ratio to be constant irrespective of vintage. If, for instance capital goods less than four years old has no depreciation, R will simply reflect this mismeasurement of capital goods of recent vintages, a mismeasurement that has nothing to do with the embodyment hypothesis.

But anyhow R has to do with mismeasurement of our capital input variable, and even if it is a bit doubtful if we should interpret a positive and significant coefficient of C as a result of embodied technical change, the situation in this case is somewhat less problematic than in the previous case where we investigated the possibilities present to trace the effects of costs of change.

iii) The effects of transitory variation in demand.

In the short run we may expect that the establishments have adjusted themselves to what they expect to be a normal demand for their goods.[13] The actual demand may show short run variation, whether purely random, or of a more systematic kind usually denoted business cycles variations. These variations, at least those of the first kind, are rather unpredictable, and they imply in general variations in the capacity utilization. Now, we have in our study no direct measure on capacity utilization or business cycles variations either for the individual establishments or even for the different industries. The question is then if any of these informations available to us show variations due to short run variation in demand, and if these variables can be used to take care of the effects of such variations. This is of course a priori an unvarranted question. But we may believe that a least two of our variables will show such variations.

---

13) Some establisherents may rather have adjusted them to a normal supply of raw materials, for instance fish and herring oil and meal factories.

These two informations are "repairs and maintenance of the capital stock" and the "inventories of finished goods". Neither of these can be assumed to be scale-independent. That is, some current repair and maintenance of the capital stock has to be carried out anyhow "to keep the wheels going" and this can be assumed to be approximately proportional to the capital stock. [14] And for the inventories of finished goods there is presumably in general proportionality with the scale of operation. But "on the margin" there is a certain degree of flexibility of both. In short-lasting recessions the establishments, instead of reducing the labour stock, can produce for inventories and/or partly use this factor of production for repairs and maintenance of the other factor of production, namely capital. And when a positive shift in demand occours the establishment can reduce the inventories of finished goods and/or reduce the repair and maintenanced of the capital goods to a minimum.

Thus we'll apply the following normalized "repairs and maintainanced variable":

$$(15) \qquad S = \frac{I_2}{K}$$

and the following normalized inventories-variable

$$(16) \qquad T = \frac{H_3}{N}$$

to take care of short run variations in demand.

As usual we can think of a number of qualifications in connection with the results of these variables too. For instance S may reflect the fact that those establishments with better maintenance of the capital stock have a higher productivity and as this isn't taken care of in our capital measure we may get a positive coefficient of S, while it should be negative according to the role we have asserted that repairs and maintenance plays, due to transitory variation in demand. A significantly positive coefficient of S may, therefore lead us to reconsider our capital measure. But if it in fact is significantly negative we have, I think, fairly good evidence of the role repairs and maintenance is assumed to play.

---

14) It does, however, probably depend on the age distribution of the capital goods, and we should possibly try to take this into consideration.

As concerns the second variable we should probably rather deflte with Y instead of N. As this is not done it is because we consider Y to be more endogeneous than N even if S is strongly related to L which probably also have an element of endogeneity. In any case, if S plays the role it is assumed to do some of the transitory variation in demand is reflected in Y and thus also in V which in most cases is our outputmeasure. Thus in this case we get a spurious correlation between lnV and $T_1 = \frac{S}{Y}$. Therefore chosen a theoretical less, but empirically presumably more, appropriate deflator, namely N.

e) The form of the production function and the problem of factor substitution.

Our first approximations to the production and behaviour relations are the value added Cobb Douglas relation and the ACMS relation respectively, both presented in section a) of this chapter. The extensions we would like to make in this context concerns mostly the production function.

First we'll try the so called Kmenta relation which is a Taylor-expansion of the CES relation arond a value of the elasticity of substitution of one, that is, the Cobb Douglas case. This relation will be tried even if the prospects of useful results are rather poor, particularly due to the poor measure we have of capital input.

Second, we'll by means of simple methods investigate if the elasticity of scale is independent of scale, particularly if it is decreasing with scale, as this, theoretically is an interesting case as it implies a u shaped marginal costs curve.

Third we'll look at our labour input measure and particularly consider if it is appropriate to have only one measure for this type of input. That is, as an alternative we split the labour-input variable into two, one concerning production workers, and one concerning employees and owners.

Fourth we'll consider more closely the role of the third factor of production, namely raw materials. We'll look at some problems of estimating the output elasticity of this factor of production and consider alternative specifications of the gross production-value production function.

Fifth we'll look at the index problem when constructing measures for output, and input of raw materials. We then have to use another body of data and this falls somewhat outside the frame of the main part of the study.

Sixth, as concerns estimation methods we'll particularly consider the possibility of applying instrumental variables and the two stage least square method.

Concerning the behaviour relation we'll first look at a simpler specification than the ACMS relation, namely the one we get when we have a Cobb-Douglas relation. We then have an elasticity of substitution equal to one. By means of this relation we are particularly interested in factor-share estimators on the elasticity of labour.

Second we'll discuss some modified versions of the ACMS-relation. We'll analyse a simple lagged adjustment model and a model where there is assumed to be first order serial correlation of the residuals, but not lagged adjustment behaviour. And in this context we'll discuss certain identification problems: What is the effect of lagged adjustment and what is the effect of auto-correlated error-terms.

Third we'll in general consider the identification problem as concerns the elasticity of substitution between labour and capital. And also, but more in the context of the investigation of production functions, discuss the identification problems as concerns the elasticity of substitution between raw materials and the value added factors of production.

Finally, we'll, if possible try to say something about the homotheticity-assumption, that is if the form of the isoquants really can be assumed to be the same, independent of scale. This question can be studied both by means of production functions and by means of behaviour relations.