AN ANALYSIS OF EFFICIENCY IN RETAILING

B. NOOTEBOOM

REPORT 7706/ES

ERASMUS UNIVERSITY ROTTERDAM, P.O. BOX 1738, ROTTERDAM, THE NETHERLANDS
AN ANALYSIS OF EFFICIENCY IN RETAILING

by B. Nooteboom

Abstract

In an earlier study a linear model of store operation was developed on the basis of queuing theory. In the current study an attempt is made to aggregate this model, by an explanation of the micro variables, associated with the instruments of store operation, in terms of macro variables, which describe the external conditions under which shops are operated, resulting in a system of simultaneous equations. The purpose of the study is to provide a basis for later empirical studies.

Contents

1. Introduction
2. Measures of productivity
3. Basic causal structure
4. Extended causal structure
5. Sales
6. Synthesis

March 1977.

The study discussed in this paper is part of a project undertaken by the Research Institute for Small and Medium-sized Business in the Netherlands (E.I.M.) in cooperation with the Econometric Institute of the Erasmus University Rotterdam.
1 INTRODUCTION

In a previous theoretical study we developed a micromodel of store operation, describing the capacity of a "production factor" (labour, shopspace) in relation to the size of sales turnover, for shops with a given assortment composition, service level and type of organisation\textsuperscript{1)}, as a function of some of the "instrumental" variables of store operation and characteristics of the environment of the shop \cite{13}.

The model was developed to provide a basis for empirical studies of "productivity" in retailing. One can conduct such a study on a micro level, with the ultimate aim of improving the performance of individual shops in their own specific conditions. It is our purpose, however, to conduct a more aggregate study, to analyse the position of a given type of shop as a whole, to provide a basis for public policy. To do this we must make the transition from our micromodel to an aggregate model to explain "productivity" in terms of macrovariables.

We have no confidence at all that we would arrive at a valid macromodel by substituting aggregated microvariables in a micromodel. Instead, we think, we should trace the causal structures linking microvariables to macrovariables. Generally, this implies a shift of attention from the instruments of micropolicy to the conditions under which those instruments are applied. This is not only better from a logical point of view, but it is also more useful, since instruments of macropolicy are seldom aggregates of the instruments of micropolicy.

The purpose of our work is to develop a general model explaining the "productivity" of any given type of shop. This gives us a model on a "meso" level, in between the micro level of individual shops and the macro level of retailing as a whole. Ultimately we want to build up a picture for retailing as a whole from the different types of shop involved, while taking into account their mutual competition.

In this report we first consider some of the conceptual problems involved in the measurement of "productivity" in retailing. Subsequently we present the micromodel, giving a basic causal structure of "productivity".

\textsuperscript{1)} such as: chain stores, voluntary chains, cooperatives, independents
This will then be developed into an extended causal structure in terms of macro variables.
The analysis is entirely theoretical and provided the basis for later empirical studies.

We are indebted to previous investigators, notably Bak [1], Bucklin [2], Mc. Clelland [3, 4], George [7], Hall c.s. [8], Holdren [10] and Palamountain [14].
MEASURES OF PRODUCTIVITY

The main conceptual problem in a study of "productivity" in retailing is that we do not and cannot have a measure of "real" output. Therefore, to begin with, we should not really use the term "productivity" at all. One might devote a separate paper to a full discussion of this problem, but it can be summarized as follows:

- The "product" of retailing is a "bundle of services"\(^1\), which presents a conceptual problem for the definition of "volume". We note that we are not only dealing with different products, each of which might by itself have a measure of volume. We are also dealing with services, and a service by definition does not have a physical measure. Therefore output can only be measured indirectly in the form of a "value", i.e. as a certain amount of money paid at a certain moment.

- On a time series basis retail price indices are available, but they generally refer to the aggregate consumption of a certain class of products rather than the sales through a specific channel of distribution. In a cross section study, differences in retail prices are generally not available at all.

The important point is, however, that even if the price differences were available, they would not always or even generally reflect "real" differences in service, due to imperfections of competition.

Therefore we cannot in general be confident that we approximate "real" output by deflating the value of output by the prices, if those were available (which they are not).

Because we cannot be confident about any attempt to measure "real" output, we prefer to use the more general term of "efficiency" instead of the more narrowly defined and more ambitious term of "productivity". This "efficiency" is what Dreesmann \(^5\) calls "commercial productivity".

In connection with this problem we note that in retailing we should not only consider the operating efficiency, as one tends to do if one considers retailing in an analogy to manufacturing. Mc. Clelland \(^3\) reminds us that in retailing commercial efficiency takes precedence over operating efficiency. Commercial efficiency refers to the way in which a shopkeeper may create and utilise opportunities in the choice of location, shop design and appeal, assortment composition, price mixing with "loss leaders"\(^2\), promotion, special services etc., in order to influence the volume, price or other aspects of turnover to his advantage.

\----

1) cf Hall c.s. \(^3\)
2) loss leaders are products with a high price elasticity, sold at reduced prices yielding a loss, to attract more customers, while the loss is recouped in other less price sensitive products.
We note that it is because competition is imperfect, in the classical sense of economic theory\(^1\), with market segmentation, spatial monopolies, product and price differentiation, demand dependence between different products in the assortment, limited information on the part of consumers, psychological and social aspects, that commercial efficiency is relevant. This is another way of repeating that neo-classical economic analysis is inadequate, as we argued elsewhere [11], and that the concept of "productivity" is problematic.

We associate operating efficiency with the actions taken to "minimise" costs at a given level and distribution over time of the volume of sales, while commercial efficiency is concerned with attempts to influence the level and distribution of sales volume, sales prices and purchase prices.

In fact, however, there is no clear boundary between the two types of efficiency. We can illustrate this as follows:

One problem in retailing is that demand varies over time, with peaks in certain periods (before Christmas, on Saturdays, during certain hours of the day) and slack demand in others (during bank holidays, on Mondays, during other hours of the day). Given a certain pattern of demand the efficiency of labour can be relatively high if one can adjust the sales capacity to the demand fluctuations by using part-time labour. Alternatively, or additionally, one can try to dissociate certain activities from the presence of the customer, so that they can be performed during the quiet hours. We call those activities "Pre- and Post Purchase Activities" (PPA). The use of part-time labour and PPA are clearly aspects of operating efficiency.

However, sales staff not only satisfy a given demand, but may also stimulate demand, particularly in the field of durables and luxuries. The use of part-time labour may then result in a more uneven, and hence sometimes inadequate, quality of sales staff, resulting in a reduced stimulus of sales. This is reported in a study of the British Economic Development Committee for the Distributive Trades [6]. Here part-time labour also becomes a matter of commercial efficiency.

The share of PPA in total labour activities that can be achieved depends on the service offered and is thereby also subject to commercial limitations. Finally, the time distribution of demand may be altered by a discrimination of prices or services, with more favorable shopping conditions during the quiet hours, and this also is clearly a matter of commercial efficiency.

\(^1\) This does not mean that competition may not be fierce.
Although the basic orientation may be different, the two types of efficiency interact to yield a given observed value of sales, gross profit and net profit. Even if we could separate the volume and "the" price of sales in our observations, which we can't, we still could not separate the effects of operating efficiency and commercial efficiency.

We can analyse the effects of factors such as demand fluctuation, PPA, part-time labour, sales turnover and prices, but we cannot say whether these are the result of operating efficiency or commercial efficiency.

In our studies we employ the concept of a "type of shop", defined as a class of shops which are homogeneous with respect to the assortment composition, the extent and nature of services offered and their contributions to the physical production and distribution of the products sold. With this definition different shops within a shoptype are homogeneous with respect to their added value per guilder of sales, and we may take sales turnover as a measure of "output".

In a comparison between shoptypes, on the other hand, one should take into account the differences in added value per guilder sales. By way of approximation one can use the gross margin, defined as the sales value minus the purchase value of turnover, as a measure of output.

Another property of the concept of shoptype is that in a study of shops within a shoptype substitution between "production factors", notably between labour and shopspace, is limited, since it would tend to imply differences in the service level, so that the shops would not belong to the same type. This justifies a separate consideration of the efficiency of labour and of shopspace, without the need to consider them jointly in the framework of some sort of production function 1).

A rigorous application of the concept of "shoptype" might yield as many shoptypes as there are individual shops. Empirically we approximate the concept by taking a "trade", which is defined according to the product group (groceries, dairy products, drugs etc.) which has the largest share in sales (grocers, milkmen, druggists), with a subdivision according to the following properties:

- Method of sale and service level. Methods of sale can be distinguished as follows: counter-service, self selection, self service, mail order, vending machine, itinerant trade. The service level is defined in the

1) in other words: apart from technical progress (which is probably biased), we are dealing with "fixed technical coefficients" if the shops are indeed sufficiently homogeneous with respect to their type.
broad sense, including the range of choice offered in the product line, deliveries, credit, and atmosphere.

- Type of organisation: chain-store, voluntary chain, cooperative, independent, department store.

The merit of this demarcation is, that it enables us to use the inter-firm comparisons of the Research Institute of Small and Medium-sized Business in the Netherlands (E.I.M.), which form the main source of our data, while it also gives an a priori reasonable approximation to the abstract concept of shop type.

In productivity studies one usually considers output per unit of a "production factor". As "production factors" we will consider labour (expressed in hours worked per annum, or in persons engaged in terms of "full-time-equivalents" (FTE), to correct for part-time labour), and shop space (expressed in square meters total shop space or selling area).

We choose labour because of its importance for public wage- and employment policy. We choose shop space for its importance in urban planning and because it seems closest to the concept of "capital".

Apart from labour and shop space there are other operating costs, such as materials (for packaging, repairs or the exercise of a handicraft), depreciation and maintenance of means of transport and of inventory, promotion expenditures, interest on working capital, insurance, light and heating, and administration. We assume that in terms of prices in a given base year those costs are constant with respect to either labour, or shop space or sales (at base year prices).
3 BASIC CAUSAL STRUCTURE

In a previous study [12] we found linear relationships between costs and turnover on a cross section basis with individual shops within a shop-type:

3.1) \[ H_{ij} = d_{1i} + e_{1i} Q_{ij} \]

3.2) \[ S_{ij} = d_{2i} + e_{2i} Q_{ij} \]

where:
- index \( i \) refers to the shop type
- "j" an individual shop (or to the average for a class of sales turnover)
- \( H = \) total hours worked per annum (per shop)
- \( Q = \) sales value of turnover per annum (per shop) ("sales" in short)
- \( S = \) surface area of the shop

The intercept (\( d_{1i} \) and \( d_{2i} \)) indicates a "threshold cost" which one incurs to operate a shop of type \( i \) at all, regardless of its sales size. In the case of labour (\( d_{1i} \)) it equals the total opening hours of the shop times the number of independent departments in the shop. The explanation of this is that someone must mind the shop (or department) whether or not there are customers present. For a more detailed discussion we refer to [13].

In general we can write:

3.3) \[ K_{si} = d_{si} + e_{si} Q_{ij} \]

where:
- the index \( s \) refers to the category of operating cost.
- \( K \) refers to a cost.

This linear structure leads to hyperbolic effects of scale:

3.4) \[ \frac{K_{si}}{Q_{ij}} = e_{si} + \frac{d_{si}}{Q_{ij}} \]

In words: the average costs of type \( s \) per guilder turnover within shop type \( i \) is equal to \( e_{si} \) plus a hyperbolic component which represents the effect of scale.

In our studies of efficiency we can thus eliminate the effect of scale by considering:

3.5) \[ P_{si} = \frac{K_{si} - d_{si}}{Q_{ij}} \]

On the average \( P_{si} \), taken over the individual shops \( j \) of type \( i \), is equal to \( e_{si} \).
We note that it is very useful to be able to separate the effect of scale from other factors in the explanation of efficiency. If one is unable to do this, one is always uncertain whether a given factor which is found to be associated with efficiency has an effect through an effect on sales size or apart from the effect of scale.

In a theoretical analysis of store operation we extended the basic cost structure into a more detailed model based on queuing theory [13]. For labour, the model is as follows:

\[ 3.6 \]

\[ H = d + \frac{1}{\varepsilon \lambda a_1} (\frac{\phi}{\pi}) a_2 \left\{ \frac{\beta_1}{tr} + \frac{\beta_2}{\psi} + \frac{1}{Q+h} \right\} Q \]

where:

- \( d \) is the threshold labour, equal to the opening hours of the shop times the number of separately staffed departments.
- \( \varepsilon \) indicates the efficiency of management.
- \( \theta \) indicates the effect of "technical factors", including "hardware" (\( \theta_1 \)) such as cash registers, trolleys, fork trucks, conveyor belts, labelling devices etc., and "software" or "know-how" (\( \theta_2 \)) such as methods for stock control and administration.
- \( \lambda \) is a measure of the quality of labour, for which we might take the relative wage rate.
- \( \phi \) is a measure of the fluctuation of demand, so that \( \phi = 1 \) when demand is constant and \( \phi > 1 \) when it is distributed unevenly over time.
- \( \pi \) is a measure of the fluctuation of labour capacity, achieved by means of part time labour, with the property that \( \pi = \phi \) when the adaptability of labour capacity is complete and \( \pi < \phi \) when it is not. As an approximation to \( \pi \) one might take the share of part time labour in total labour.
- \( \beta_1 \) is the average time spent per customer.
- \( tr \) is the "transaction size", defined as the average sales per customer (per customer visit, to be more precise).
- \( \beta_2 \) is the average time spent per customer per guilder sales.
- \( Y_2 \) is the average time spent on PPA per guilder sales.
- \( \psi \) indicates the share of wholesalers and producers in distributive activities such as transporting goods to the retailer, breaking bulk, storage, packaging, price labelling, display etc. We call these "purchasing economies" in short.
is the average waiting time of customers as a ratio to the average service time, called the "relative waiting time".

\(h\) is a parameter associated with the shape of the distribution of labour requirements over time, as a function of the sales size.

\(Q\) is annual turnover (sales size).

The dimensions of these parameters are as follows:

- \(\varepsilon, \theta, \lambda, \phi, \pi, \psi, w\) : dimensionless
- \(H, d, \beta_1\) : hours
- \(\beta_2, \gamma_2\) : hours/guilders
- \(tr, h, Q\) : guilders

If we assign fixed values not only to \(\beta, \beta_2\) and \(\gamma_2\), but also to \(\varepsilon, \theta, \lambda, \pi, tr, \psi\) and \(h\), for all shops (index \(j\)), we can give a graphical representation of the values for \(H\) as a function of \(Q\) for a given range of values for \(w\).

This is the cigar-shaped area in figure 3.1.

\[
H_{ij}
\]

\[
Q_{ij}
\]

In the empirical studies we envisage, we expect to have no observations on the relative waiting times \(w\).

We define:

\[
x_{ij} = \frac{\frac{1}{w_{ij}} - 1}{\frac{Q_{ij}}{Q_{ij} + h_{ij}} d_i}, \text{ which is a term of 3.6}
\]
Having no observations on \( w_{ij} \), we would like to ignore \( X_{ij} \) in 3.6) and consider it as part of a general disturbance term. This will not introduce any bias, provided that the expectation of \( X_{ij} \) is zero, which is plausible when the average value of \( w_{ij} \) is 1.0. This implies that the average waiting time of customers is, on the whole, about equal to the average service time. This may be plausible, but we do not know if it is the case, for retailing in general. Suppose it is not true. Taking an extreme case, suppose the value 1.0 lies outside the range of actual values of \( w \). Then the shaded area of figure 3.1 will no longer be symmetric and will turn up or down when \( Q \) approaches zero.

In figure 3.2 we illustrate the case that the actual values for \( w \) are larger than 1.0.

If such a situation occurs, there are two possibilities for our empirical results:

a. If the observations include many relatively small shops, the cost curve will be non-linear. This might be inferred from the Durbin Watson ratio in a linear regression of \( H \) on \( Q \).

b. If the observations include only the relatively large shops, a linear regression of \( H \) on \( Q \) will yield an intercept which is, in the case of figure 3.2, appreciably lower than \( \delta_{ij} \).
When we drop $X_{ij}$ from 3.6), on the assumption discussed above, the model becomes:

$$H = \delta + \frac{1}{\epsilon \theta \lambda^{a_1}} \left( \frac{\hat{u}}{m} \right)^{a_2} \left\{ \frac{b_1}{r} + b_2 \frac{\gamma_2}{\psi} \right\} Q + \delta$$

where $\delta$ is a stochastic disturbance term.

So far, we have concentrated on labour, but we must also consider the shop space. Does the same model structure apply?

Firstly, there is a threshold in shop space as well, associated with a minimum range and volume of products on sale in order to belong to a given shop type. It is not, however, associated with the opening time of the shop, as the threshold labour is, so that it varies more between shop types and will not become smaller when the opening time is reduced.

Secondly, we should ask whether the results based on queuing theory apply to shop space as well. This question may require further theoretical analysis, but at this stage we see no evident reason why the theory should not apply.

The space occupied by a sales attendant and sales facilities is as much part of the service channel as the attendant himself; to reduce average waiting time at a given level of demand one must add shop space in much the same way that one must add labour capacity, and the utilisation of shop space increases with the number of service channels, and hence with sales size, in the same way as the utilisation of labour does. This argument applies to the space associated with serving customers. In self-service stores, however, space is also used differently for the customers to perform self selection and self service, but we will not analyse that further at this stage.

Thirdly, we note that there obviously is no such thing as part-time shop space. Although that does not fundamentally alter the model structure, there is the following consequence: demand fluctuations cannot be absorbed by an adjustment of capacity and therefore have a direct effect on "floor productivity". In terms of our model: $\pi = 1$ in 3.6).

Fourthly, it is plausible that a certain fixed amount of space capacity is required per customer ($b_1$), a certain variable amount per guilder sales ($b_2$), and a certain amount per guilder sales for PPA ($\gamma_2$). In self-service stores $\gamma_2$ will be relatively low: the space required for handling (storing) goods to some extent coincides with the space needed for customers to make their purchases. $\gamma_2$ is associated with storage space off the sales area, office space, rest-rooms etc. In fact, one might even use the ratio between sales
area and total business area as an indicator of the extent of self service. This can be misleading, however, since storage space will be smaller not only when there is more self service, but also when there is a more frequent supply from the wholesaler, as part of "purchasing economies".

Finally we note that:

- $\lambda$ in 3.7) now serves as an indicator of the "quality" of the shop. Even more than in the case of labour, however, the quality of the shop, associated with the quality of the location, is much more a matter of commercial than of operating efficiency. It then reveals itself mostly in other variables such as sales size ($Q$), including the price level of sales, the transaction size ($tr$) and the (lack of) demand fluctuation ($\phi$). A priori it is doubtful that the quality of the location is reliably expressed in the rental price or construction cost of the store.

- the "soft ware" part of the technical factors is now associated with the design of isles, shelves, counters, check-outs etc. Here also the operating and commercial aspects are intertwined.
EXTENDED CAUSAL STRUCTURE

In the following analysis we concentrate on labour efficiency. In our micromodel we specified a number of "causes" of the efficiency of labour. These causes are of different types. It is useful to employ the Aristotelian categories of "efficient causes" reflecting the action of people, "material" causes reflecting the "instruments" used and "conditional" causes reflecting the conditions under which people use the instruments. The "efficient" causes are management efficiency (ε) and labour quality (λ). The "material" causes are the technical factors (θ), part-time labour (π), purchasing economies (ψ) and the relative waiting time (w). Subject to the constraints of assortment composition and service level within a shop type, the service time (β1, β2) and the time required for PPA (γ2) are also material causes. The fluctuation of demand (φ) the transaction size (tr) and the sales size (q) are largely "conditional" causes, depending on the characteristics of the location of the shop.

These distinctions indicate a difference in emphasis rather than any sharp boundaries. The efficient and material causes are to some extent constrained, and in that sense are like conditional causes: management efficiency depends on competition, the labour quality and the available part-time labour depend on labour market conditions, the technical factors depend on technical progress, the purchasing economies that can be achieved depend on the organisational structure of wholesalers and producers, the relative waiting time that one can afford depends on consumer preferences and on competition, and so does the service time. On the other hand, the location of the shop is an instrument "in the long run", and at a given location one can affect the sales size, the transaction size and perhaps even the demand fluctuation.

On the whole, the emphasis is on specific instruments of store operation, and that is what makes the model a micro model. To put this differently, the causes are "close", logically speaking, to store operation. Following Hall c.s. we call them the "proximate causes" of efficiency [8].

In an empirical micro study one would try to specify observable variables which represent the proximate causes as closely as possible. The ultimate aim of such a study is to improve the operation of individual shops. It is our task, however, to provide a basis for the determination of public policy, and the instruments of government (be it local or on the state level) are not the same as the instruments of the shopkeeper. The instruments of public policy are associated with the conditions under which shops are operated, and are "further away", logically speaking, from the instruments of store operation. In a macro model, therefore, the emphasis will shift
away from the material causes to the conditional causes; the conditions under which shopkeepers select their instruments. We will call these "remote causes", reflected in macro variables such as consumer income, wage rates, employment, competition, prices, spatial structure of retailing outlets etc.

Ideally, we should perhaps construct a system of simultaneous equations in which the proximate causes are specified as functions of those remote causes for which "meaningful" macro variables can be found. "Meaningful" means here, that the macro variables should:
- give as close a representation of remote causes as possible
- provide a link with instruments of public policy
- provide a link with other macro models and with official statistics
- be observable (empirically specifiable)

To specify such relationships we must trace the causal structure, connecting the proximate and remote causes. The most systematic way of doing this is probably to explain each of the parameters in the micro model as a function of macro variables. We will attempt to do this for the labour costs, but first we will briefly discuss what we mean by a "variable" and how variables are used in our study.

One type of variable, arising out of theoretical considerations, may be called a "variable" purely because it appears as such in a mathematical function, i.e. it appears in some causal structure, but the question whether it has an empirical correlate has not yet been asked. "Management efficiency (e)" is an example of such a variable.

For another type of variable it seems clear that an empirical correlate can be found. It is a "true" variable in the sense that it may be observed and may in that sense be said to "exist". It may still be rather imprecise, like a word which may have different meanings in different contexts, and which may have more or less close synonyms, but it allows for the prescription of an observational procedure, which always depends on a specific operational context. "Demand fluctuation (f)" is an example.

A third type of variable is associated with a specific observational procedure such as: "Demand fluctuation measured by such and such a procedure". (For example: "If g(t,Q) gives the distribution of the rate of customer arrivals over time t, in the course of a year, at a total annual turnover Q, the demand fluctuation is measured by the integral of g(t,Q) over values of t for which g(t,Q) exceeds the annual average rate of arrivals, divided by the integral of g(t,Q) over all values of t").
These three types of variables are close to Haavelmo's concepts of "theoretical variables", "true variables" which allow for an "experimental design", and "observed variables", and we will adopt that terminology. As Haavelmo notes, the observed variables may not satisfy the "experimental design", and any discrepancy must be watched carefully: if it becomes too large one may have to revise one's theory.

We would not say that in the construction of a model each variable should at all stages be tested for the existence of an empirical correlate: A theoretical variable without an empirical correlate may fulfill a legitimate purpose in a theoretical study. But for empirical application a model must ultimately be developed to a stage where the variables are all "true". In practice one may have to go further and specify the model in terms of variables which not only allow for observation but for which observed variables are already available. These may not satisfy the observation procedures required to bring us close to the true variables. The question then is whether the available data provide either "proxy" variables to the unobserved variables, or whether those can in turn be explained by the available data in an extended model. In the case of a proxy variable we are not dealing with a variable which does not quite satisfy the observation procedure associated with the desired variable, but one which does not satisfy it at all but is expected to be closely correlated with that variable. In the case of an extended model we postulate that the unobserved variable is some function of observed variables. In either case we solve our problem on the basis of assumptions which cannot be tested separately.

For example, let us consider "management efficiency" (ε), and ask how it is to be observed. It may be observed as a "residual" in the efficiency of labour after we have accounted for the effects of the other explanatory variables. But then we are going around in a circle: we wanted to observe ε to contribute further to an explanation of labour efficiency. A solution might be to postulate that ε depends on the person of the shopkeeper, notably his age, experience and training, for which observations may be available. The general validity of this hypothesis might be tested on basis of the observed residual of labour efficiency, after elimination of the effects of other variables, but in a different operational context, with other data than the ones we are using in the study at hand.

In our current study we are not concerned with the problem how to give an

---

1) The discussion gives our personal interpretation of Haavelmo's concepts.
empirical specification of the variables in the micro model but with the empirical specification of the macro variables which appear as explanatory variables in the relationships postulated to explain the micro variables. For a direct test of such a relationship one still needs an empirical specification of the micro variables. Such tests might provide the subject for further micro studies. In our study we will concentrate on indirect tests of the ultimate effects of the remote causes (macro variables) on the observed labour efficiency, without testing the intermediate effects of the remote causes on the proximate causes (micro variables) and of the proximate causes on labour efficiency. Admittedly, such "indirect testing" may provide results of which the interpretation is ambiguous, as we shall see. After this introductory discussion we will now proceed to the analysis of causal structures in terms of remote causes.

First we consider the "efficient causes".

a. The management efficiency (ε) depends to some extent on the type of organisation, and hence on the shotype, in the sense that it is possibly greater and certainly more even for chain stores than for independents. For independents it depends on the level of (un)employment, to the extent that during a depression people may "take refuge" in shopkeeping, leading to an increase of the number of "subsistence shops" and a lower average level of management efficiency. In view of contemporary social security and of increasing "professionalization" of retailing this effect is now probably less than it used to be. ε may further depend on age, education and experience of the shopkeeper. Hence:

\[ \epsilon = f(PVU, TO, AS, ES) \]

where:

PVU is the percentage of vacancies minus the percentage of unemployed on the local labour market\(^2\)

TO is the time that the shopkeeper has been in business (as a measure of his experience)

AS is the age of the shopkeeper

ES is the number of years of education/training

With plus or minus signs above the explanatory variable we indicate whether its effect is positive or negative. Of the age of the shopkeeper it is difficult a priori to say what its effect will be.

\[ \text{(4.1)} \]

1) Shops yielding a mere subsistence income, with insufficient capital, experience and perhaps motivation to flourish.

2) This variable was introduced and used empirically by George \([7]\).
b. As a measure of the quality of labour (\( \lambda \)) we might take the wage rate \( L \), so that:

\[
\lambda = f(L)
\]

The implicit assumption of this specification is that either or both of the following conditions apply:
- the differences in the official negotiated wage scales, which are largely based on seniority, actually reflect differences in quality.
- in practice employers stretch the official scales to reward individual differences in "productivity".

Secondly, we consider the "material causes". To begin with, we consider those which are closely associated with the properties of the shop type and hence should show only a limited variation in the study of a given shop type.

c. Threshold labour \( (d) \) satisfies:

\[
d = n \cdot s\]

where: \( n \) is the number of "departments", defined as sections of the shop for which at least one person should be available throughout shopping hours.

\( s \) is the annual opening time of the shop.

d. The service time per customer \( (\beta_1) \), the service time per guilder sales \( (\beta_2) \) and the PPA per guilder sales \( (\gamma_2) \) are associated primarily with the service level and the types of products sold, which are properties of the shop type. Within a shop type, however, there may be differences in the "depth" of the assortment\(^1\) and in the smaller details of service. We expect more well-to-do customers to demand more service and choice. When competition is intense, this may also increase both the service level and the depth of the assortment. When labour is expensive and/or scarce we expect labour saving to result in a reduction of service. Therefore we specify:

\[
\beta_1, \beta_2, \gamma_2 = f(I, \text{COM}, L, \text{PVU})
\]

where:
- \( I \) is the per capita income of consumers
- \( \text{COM} \) is a measure of the intensity of competition
- \( L \) is the wage rate
- \( \text{PVU} \) is a measure of the scarcity of labour, as defined before

\(^{1}\) defined as the number of different qualities and brands for each type of product
e. The opportunities to use technical factors ($\theta$) depend partly on the
assortment composition and service level, partly on the type of
organisation, and hence on the type of shop. For a given type of shop
it will depend on "technical progress" and hence on time ($t$). The extent
to which the available opportunities are used probably depends on the
intensity of competition.
Hence:

$$\theta = f(t, \text{COM})$$

f. The opportunities to obtain purchasing economies ($\psi$) will depend on the
assortment composition, the service level and on the bargaining position
of the retailer, and hence on the type of organisation, and hence $\psi$ depends
mainly on the shop type.
The opportunities will also depend on developments in the organisational
structure of wholesalers, producers and retailers taken together, and
hence on time. Within a shop type, and in a certain year, it may yet differ
between shops (depending on the frequency of supply and hence the average
stock size and on who takes care of the physical distribution). Again,
intense competition may stimulate economies. Hence:

$$\psi = f(t, \text{COM})$$
The sign of the time trend is not clear a priori.

The remaining material causes ($\pi$ and $w$) are not, or at least much less,
associated with the shop type:

g. part-time labour ($\pi$) may be observed directly, in the form of the share
of part-time labour in total labour (SPL), which might also be a meaning-
ful macro variable. In terms of remote causes it will depend on the need
to save labour, either due to the pressure of costs, particularly the
cost of labour ($L$), or due to the scarcity of labour. Hence:

$$\pi = f(L, \text{PVU})$$

h. The relative waiting time ($w$) will depend on the demand for quick service
and hence, we assume, on income ($I$), and on the intensity of competition.
It may also be raised as a consequence of the high cost or scarcity of
labour. Hence:

$$w = f(I, \text{COM}, L, \text{PVU})$$

Thirdly, we are now left with the conditional causes: $tr$ (transaction size),
$\phi$ (demand fluctuation) and $Q$ (sales size). To specify functions for these
variables we have to consider the characteristics of consumers and the
spatial structure of retailing outlets, to which we dedicate a separate
paragraph.
Before we proceed we will bring the results of this paragraph together in a simplified form.

As an intermediate step we define:

1. A "generalised measure of the service level" (GSL), which depends on the service level \((\beta_1, \beta_2, \gamma_2)\) and on the relative waiting time \((w)\), as follows:

\[
4.9) \quad \text{GSL} = f(\beta_1, \beta_2, \gamma_2, w)
\]

2. A "generalised measure of efficiency" (GEF) which depends on management efficiency \((\varepsilon)\), labour quality \((\lambda)\), the use of part-time labour \((\pi)\), technical factors \((\theta)\) and purchasing economies \((\psi)\), as follows:

\[
4.10) \quad \text{GEF} = f(\varepsilon, \lambda, \pi, \theta, \psi)
\]

According to the basic linear cost structure we had:

3.1) \( H = d + eQ \)

According to the micro model of store operation we have from 3.6):

\[
4.11) \quad e = f(\varepsilon, \theta, \lambda, \phi, \pi, \beta_1, \beta_2, \gamma_2, \text{tr}, \psi, w)
\]

Using \(4.9)\) and \(4.10)\) we can specify this as follows:

\[
4.12) \quad e = f(\text{GSL, GEF, } \phi, \text{tr})
\]

From \(4.4)\) and \(4.8)\) and \(4.9)\) we find:

\[
4.13) \quad \text{GSL} = f(I, \text{COM, L, PVU})
\]

From \(4.1)\), \(4.2)\), \(4.5)\), \(4.6)\), \(4.7)\) and \(4.10)\) we find:

\[
4.14) \quad \text{GEF} = f(\text{COM, L, PVU})
\]

Here we have left out the details in \(4.1)\) concerning the background of the shopkeeper. The reason is that we do not in general expect them to be available in the data.

The simplified "system" of \(4.12)\), \(4.13)\) and \(4.14)\) is illustrated in the causal diagram of figure \(4.1)\).
The service level (GSL), which has a negative effect on labour efficiency \((1/e)\), will be higher at a higher level of consumer income \((I)\) and at a greater intensity of competition \((COM)\), but when labour is expensive \((L)\) or scarce \((PVU)\) this will lead to labour saving through a reduction of service.

The efficiency with which the store is operated \((GEF)\) will be higher under the pressure of competition and when labour is expensive or scarce.

We note that the effects of the high cost or scarcity of labour reinforce each other: efficiency is increased and service is reduced. The effects of competition, however, are contrary and may therefore cancel out: it tends to reduce costs by a stimulus of operating efficiency but it also tends to increase costs by non-price competition in the service level.

\text{figure 4.1}
We will now consider the sales size \( Q \), the transaction size \( tr \) and the fluctuation of demand \( \phi \).

By definition:

\[
\begin{align*}
5.1) \quad Q_{ijr} &= Q_{rk} \sigma_{ijrk} \\
\text{where:} \\
Q_{ijr} &\quad \text{is the annual sales of product class } r \text{ in a shop } j \text{ belonging to} \\
& \quad \text{shoptype } i. \\
Q_{rk} &\quad \text{is the annual sales of product class } r \text{ in shopping centre } k \text{ where} \\
& \quad \text{the shop is located.} \\
\sigma_{ijrk} &\quad \text{is the share in the sales in shopping centre } k \text{ of product class } r \\
& \quad \text{achieved by shop } j \text{ (of shoptype } i). \\
\end{align*}
\]

We can say that \( Q_{rk} \) is the "local market" for product class \( r \) and \( \sigma_{ijrk} \) the "market share" of shop \( j \). The total turnover of shop \( j \) is, by definition:

\[
5.2) \quad Q_{ij} = \sum_r Q_{ijr}
\]

In our macro studies we ultimately have to explain the average sales size (and preferably also the distribution of sales sizes) for shoptype \( i \) \( Q_i \) as a function of the total consumption of product classes \( Q_r \), the market shares of shoptype \( i \) \( \sigma_{ir} \) and the number of shops in shoptype \( i \) \( N_i \):

\[
5.3) \quad Q_i = \frac{\sum_r Q_r \sigma_{ir}}{N_i}
\]

The main difference between 5.1) and 5.3) is that 5.1) adds a regional distribution of demand across shopping centres, and it allows for different market shares of shoptype \( i \) in different shopping centres.

For \( Q_{rk} \) we propose:

\[
5.4) \quad Q_{rk} = \frac{PD_r CH_r}{DE_r} \delta_{rk}
\]

where:

PD \quad \text{is the (average) population density in the region } R \text{ where the shopping} \\
& \quad \text{centre is located.} \\
CH_r &\quad \text{is the (average) per capita consumption in region } R \text{ of product class } r.

\[\text{---}\]

1) We consider a single shop as a "special case" of a shopping centre.
is the (average) number of shopping centres selling product class \( r \) per unit of area in region \( R \). Apart from a multiplicative constant 
\( 1/DE_r \) is a measure of the (average) distance of consumers to the nearest point of sale selling product class \( r \), if we assume a homogeneous spatial distribution of population and shopping centres and we ignore the spatial dimensions of the shopping centre itself. \( \delta_{rk} \) incorporates any remaining effects such as the relative attraction of centre \( k \) compared to other centres.

If we assume \( \delta_{rk} = 1 \) for all \( k \), and we assume that PD, \( CH_r \) and \( DE_r \) are homogeneous over the region \( R \), then 5.4 simply states that the total consumption of product class \( r \) is shared equally among all centres providing that product class. To give the model more content we should specify a function to explain \( \delta_{rk} \).

Even if all goods of every class \( r \) have the same price, quality and choice wherever they are sold, and PD, \( CH_r \) and \( DE_r \) are homogeneous over \( R \), \( \delta_{rk} \) will not be equal to unity for all \( k \) when there is a "hierarchy" of shopping centres selling all kinds of durables and specialty goods. In a nutshell, this is due to the fact that when a consumer travels far to purchase a durable or luxury good she may at the same time buy necessities (particularly if she goes by car) even though they could be purchased nearer by.

A model which takes this effect into account already becomes fairly complicated, let alone models which also take into account differences in price, quality and choice within each product class and in homogeneities in income, age distribution, car ownership, facilities for parking etc. To keep the current study within bounds we propose the following: \( \delta_{rk} \) is an increasing function of the total number of retailing facilities offered in centre \( k \), which we might measure by the number of shops \((N_k)\) and of the number of shops selling product class \( r \) \((N_{rk})\). In so far as the level of competition affects the price and service level this also will affect the purchases made. Hence we specify:

\[
5.5 \quad \delta_{rk} = f(N_k, N_{rk}, \text{COM})
\]

\( CH_r \) is subject to a consumption function:

\[
\text{---}
\]

1) or the total sales area of all shops, or same combined measure of sales area, number of shops and number of types of shop.
\[ C_H^r = f(I, P_r) \]

where:

- \( I \) is the (average) per capita income in \( R \).
- \( P_r \) is the price of product class \( r \) relative to the price level of some appropriate package of products, as an average over \( R \).

If we can assume homogeneous "market prices", \( P_r \) may be omitted in a cross section study.

To simplify the analysis, let us assume that shop \( j \) sells only one product class \( r \). With 5.2) and 5.1) we then have:

\[ q_{ij} = q_{ijr} = q_{rk} \sigma_{ijrk} \]

\( \sigma_{ijrk} \) will depend on the consumers' preferences for shoptype \( i \) when purchasing product class \( r \) (intertype competition) and on the number of shops of type \( i \) in centre \( k \) (\( N_{ik} \)). At this stage we cannot conduct a detailed study of preferences for different types of shop. This will later be done on a more aggregate level, in a separate study. Here we only propose that the share of shoptype \( i \) is a decreasing function of the number of shoptypes available in \( k \). When we weigh each shoptype with the number of shops available we arrive at the following specification:

\[ \sigma_{ijrk} = f(N_{rk}) \]

where \( N_{rk} \) is simply the number of shops (in centre \( k \)) selling product group \( r \).

From 5.5) we see that a large value for \( N_{rk} \) will increase the attraction of the centre with respect to product class \( r \), and hence the share of the centre in total regional consumption, but from 5.8) we see that it will reduce the share of an individual shop in that demand. The net effect of \( N_{rk} \) on \( q_j \) will probably be negative: what one loses to competitors is probably larger than what one gains together. Combining 5.7), 5.4), 5.5), 5.6) and 5.8) we then have:

\[ q_{ij} = f(I, PD, DE_r, COM, N_k, N_{rk}) \]

Now that we do not intend to explain the different preferences between different shoptypes selling product class \( r \) we may as well drop the subscript \( i \) indicating the type of shop.

The causal structure is illustrated in figure 5.1.
Now we consider the transaction size in shop $j$ ($tr$). We again assume that the shop sells only products of class $r$. We note that the customers are households rather than consumers, since one person will normally do the shopping for one household. We define $F_r$ as the average frequency with which consumers buy products of class $r$. When a customer (a household) buys products of class $r$, the volume bought each time will then, on average, be equal to:

$$5.10 \quad tr = \frac{CH_r \cdot HS}{F_r}$$

where: $tr$ is the average amount bought per shopping trip of product class $r$. $HS$ is the average household size.

Again, we should consider whether the amount bought per trip depends on where it is bought. When a housewife goes to the nearest town to buy specialty foods she may at the same time do her shopping for groceries there. Will she buy more or less than she usually does at her neighborhood grocery? Among other things, this will depend on whether she goes by car. What happens when she travels a longer distance to select an evening dress? Will she still buy groceries on the same trip or will she now utilize the opportunity to buy other more specialized goods unavailable in the nearby town, without wasting time and carrying capacity on things as uninteresting as groceries? It seems plausible that shoppers will concentrate on those products which are closest in their luxury/specialty nature to the product class which constituted the
primary goal for making the trip. We might call this a "slip stream" effect: close products are dragged along.

As indicated before, we could develop a model to take this type of effect into account, but to keep the matter simple we assume the following: The main point is the decision whether or not to buy products of a given class on a given shopping trip. Once a decision to buy is made, the amount purchased is more or less standard. We say "more or less" because we must allow for impulse buying. We might suppose that the amount bought depends on the variety offered and hence, again, on the number of shops and, again, on the relative intensity of competition in the centre.

Thus we specify:

\[ t_{rk} = \frac{CH_{r,H}S}{F_r} u_{rk} \]

where: \( t_{rk} \) is the amount bought of product class \( r \) in shopping centre \( k \).

\[ u_{rk} = f(COM, N_{rk}) \]

Now, is the transaction size in a given shop \( t_{j} \) the same as the amount purchased per visit to the centre? It will be if the customer buys the entire amount in one shop, if she buys it at all. But this amount may be distributed over several shops, and we will assume that the chance that this will happen depends on the number of shops selling the product class:

\[ t_{j} = t_{rk} v_{jrk} , v_{jrk} \leq 1 \]

\[ v_{jrk} = f(N_{rk}) \]

Combining 5.13), 5.12), 5.11), 5.12) we now have:

\[ t_{j} = f(CH_{r,H}, HS, F_r, COM, N_{rk}) \]

We see that the variety of supply in the centre \( N_{rk} \) may have both a negative and a positive effect on the transaction size per shop, and a priori it is not clear which effect will prevail.

Now we must try to explain \( F_r \).

One could study the long-term development of the density of retailing facilities \( DE_r \) as a function of consumer preferences for short travelling distances, for infrequent shopping trips and for a diversity of supply, which will depend on income, car ownership, traffic congestion, the availability of parking space and the availability of free time.
In this study, however, we are interested in the distribution of sales at a given density of outlets, and therefore we consider $DE_r$ as given (exogenous).

The frequency with which customers purchase products of a given class ($F_r$) will depend on the average distance to centres selling that product class ($1/DE_r$), on the mobility of customers, the availability of a car to transport the larger bulk of purchases which is associated with a lower frequency, the availability of domestic storage facilities to stock that bulk, and the free time available for shopping. Hence:

$$5.16 \quad F_r = f(DE_r, BT, DS, MOB, AT)$$

where:

- **BT** indicates the capacity for "bulk transport"
- **DS** " " " domestic storage
- **MOB** " " mobility
- **AT** " " available time.

The capacity for bulk transport will be less for the elderly and more for car owners, but for the latter it will then depend on the effective use of that car, which depends on parking space, traffic congestion and, at some point, on the costs of running the car:

$$5.17 \quad BT = f(A, CU)$$

where:

- **A** is the average age of the population
- **CU** is effective car usage, with:

$$5.18 \quad CU = f(CO, PARK, TC, CCU)$$

where:

- **CO** is the average number of cars per household
- **PARK** is the availability of parking space
- **TC** indicates traffic congestion
- **CCU** are the costs of car usage.

Mobility is low for the elderly, and for the rest it will depend on effective car usage:

$$5.19 \quad MOB = f(CU, A)$$

1) Or, preferably, the percentage of people above a certain age.
The time available for shopping will depend on the number of labour hours per week (WT) and the percentage of working women (WW):

\[
A_T = f(WT, WW)
\]

The overall causal structure related to \( F_T \) is illustrated in figure 5.2.

Figure 5.2 shows the predicament of the elderly: due to low mobility one does not want to shop frequently, but due to the inability to carry a large bulk one must. Hence the overall effect on \( F_T \) is probably positive. Car usage (CU) also has potentially contrary effects: an increased frequency through increased mobility, but also the opportunity to buy less frequently through the capacity to transport bulk, provided there is the capacity to store it at home (DS). We should also take into account that car usage may also increase traffic congestion, unless it is offset by road construction, and it may reduce the availability of parking space. This situation is the most likely in urban areas, and there we would expect the net effect of car usage (and car ownership) on the frequency of purchases to be negative. However, the net effect of car ownership depends on the nature of the product class (r). For basic foods the net effect is
probably negative, because of the bulk involved and the unattractiveness
of the shopping activity itself. For some luxuries/durables the bulk is
small or negligible (jewellery, books, records, clothes), and the shopping
activity itself may yield a positive recreational utility, with a positive
net effect on purchase frequency. For other durables (refrigerators, furni-
ture) the availability of a delivery service will eliminate the effect of
BT, but the net effect of car ownership is still zero because one does not
stock that kind of product at home.
Our conclusion is now as follows:

\[ 5.21 \quad F_p = f(D_t, A, CO, U, DS, WT, WW) \]

where: \( U \) is a measure of urbanisation, which has a negative effect on
shopping frequency to the extent that it is associated with less
parking space and a high level of traffic congestion.
The sign of \( CO \) depends on the nature of the product class.

Combining 5.15), 5.6) and 5.21) we now have:

\[ 5.22 \quad tr_j = f(I, HS, DE_t, A, CO, U, DS, WT, WW, N_{rk}) \]

Now we are left with the demand fluctuation \( (\phi) \).
It is useful to distinguish between the fluctuation during the day \( (\phi_d) \) and
the fluctuation during the week \( (\phi_w) \).
The second kind is more easily solved by means of part-time labour than the
first: it is easier to obtain additional staff only for saturdays than only
for certain hours of the day. Hence we should also distinguish between
hourly part-timers \( (\pi_d) \) and daily part-timers \( (\pi_w) \).
When the immediate neighborhood (within acceptable walking distance) is a
residential rather than a business quarter, people may "pop in" at odd hours
of the day. This might be measured by the share of homes in the total built
area in the neighborhood \( (SH) \).
What is the effect of the distance travelled by customers who do not live
"a few blocks away"? They will visit the shop when they have the time re-
quired to travel the distance involved. To the extent that people tend to
be "off duty" at the same moments, distance forces people to come in peaks.
However, when the area from which people come increases, there will be an
increased dispersion of travelling times, causing the peaks to last longer
and thus be flatter. Thus it is conceivable that the fluctuation during the
day first rises and then falls with the size of the area from which people
come.
There is a limit to the distance people can travel for shopping during weekdays, depending on the opening hours of shops. At distances beyond this limit people will have to come on free days, and there will be an increasing emphasis on the peaks during the week (saturdays). In the transition from daily to weekly peaks there is, as noted, a transition from difficult to more flexible part-time labour (πd to πw). We note, however, that the problems of the utilisation of the shop space may well increase.

As the area increases even further, in extremely sparsely populated areas, where shopping becomes a major expedition requiring several days, the weekly peaks are likely to subside.

Summing up, there are two dimensions to the problem: distance, modified of course by car ownership, and available time.

The size of the area served is a decreasing function of the spatial density of shopping facilities (DEr). Distance is bound to have an effect, but it is not a simple one and requires further empirical study. In a general specification of φ we cannot specify whether the effect of DEr will be positive or negative, and hence we will not include it.

When people have more time available (AT) there is likely to be less fluctuation of demand. In particular, it will be less when people have time at different moments. This will be the case for housewives more than for working women, for farmers more than for office workers, and for unemployed more than for employed.

Thus we now specify:

\[ 5.23) \quad \phi = f(SH, AT) \]

Any effect of DEr will have to be considered in a more ad hoc manner considering the characteristics of the empirical situation in hand.

With 5.20) this gives:

\[ 5.24) \quad \phi = f(SH, WT, WW) \]

---

1) The longer opening times in the U.S. and, to some extent, in Canada are probably associated with the longer distances travelled.
SYNTHESIS

We will now collect the results to form an overall system of equations.

The basic linear cost structure gave us:

3.1) \( H = d + e Q \)

From paragraph 4 we have:

4.12) \( e = f(GSL, GEF, \phi, tr) \)

4.13) \( GSL = f(I, COM, L, PVU) \)

4.14) \( GEF = f(COM, L, PVU) \)

From paragraph 5 we have:

5.9) \( Q = f(I, PD, DE_r, COM, N_r, N_{rk}) \)

5.22) \( tr = f(I, HS, DE_r, A, CO, U, DS, WT, WW, N_{rk}) \)

5.24) \( \phi = f(SH, WT, WW) \)

This constitutes a system of 7 equations in the following 7 endogenous variables: \( H, e, Q, GSL, GEF, tr, \phi \).

We cannot apply this system empirically in the studies we envisage, because we do not have observations on GSL, GEF and \( \phi \).

We can try to resolve this by reducing 4.12), 4.13), 4.14), 5.22) and 5.24) to a single equation for \( e \), but then some of the most important exogenous variables have contrary effects, which may cancel out, as we show below. We then obtain:

6.1) \( e = f(I, COM, L, PVU, HS, DE_r, N_{rk}, A, CO, U, DS, WT, WW, SH) \)

The variables with contrary effects are \( I, COM, N_{rk}, CO, WT \) and \( WW \). The contrary effects of \( COM \) were discussed in paragraph 4.

The effect of income \( (I) \) on \( e \) may be negative due to a positive effect on the transaction size and positive due to a greater demand for service.

The contrary effects of \( N_{rk} \) were discussed in paragraph 5.

The time available for shopping, depending on the duration of the labour week \( (WT) \) and on the percentage of working women \( (WW) \) may have a negative effect on \( e \) due to a smaller demand fluctuation and a positive effect due to a smaller transaction size.
We cannot say that the contrary effects will cancel out, but only that they may do so. If the elasticity of service with respect to the intensity of competition is small, competition will have a negative effect on $e$ (due to higher efficiency). Likewise if the elasticity of service with respect to income is small, as may be the case for convenience goods such as groceries (where the elasticity may even be negative due to a preference for speedy shopping), income will have a negative effect. Empirically we may therefore test what the net effect is, but a priori the effects are ambiguous.

We can try to improve the situation by looking for new exogenous variables which do not have contrary effects, and substitute those for exogenous variables which do.

Hall c.s. [8] proposed the concept of "dynamism" as an explanatory factor. It is associated with new and developing settlements, measured by the population growth (PG) and/or the age of the settlement (AS). In such an environment the newer premises might allow for better methods of store operation ($\theta$), management might be better ($\varepsilon$), wholesalers might provide favorable supply conditions to penetrate the new market ($\psi$). Furthermore, the supply of retailing outlets may lag behind the population growth and hence demand, leading to monopolistic positions allowing for longer waiting times ($w$) and less service ($\beta_1, \beta_2, \gamma_2$).

We define:

$$6.2) \quad DYN = f(PG, AS)$$

where: DYN indicates the level of "dynamism".

The effect on $e$ is illustrated in figure 6.1.
This concept of dynamism appears quite attractive, without contrary and only with mutually reinforcing effects.

Eliminating the "ambiguous" and the "minor" variables we now have:

\[ e = f(L, PVU, DYN, HS, DE_r, A, U) \]

We omitted SH as a variable which is probably of minor importance, and DS on the assumption that every household now has a refrigerator.

We note that U is actually a proxy variable for traffic congestion and limited parking space. One proximate cause which was eliminated but which may be a useful macro variable is the part time labour \((\pi)\) which might be measured by the share of part-time labour in total labour \((SPL)\) and introduced in 6.3), with a negative effect on \(e\). We note, however, that it is a "material cause" associated with the "conditional causes" of high cost \((L)\) or scarcity \((PVU)\) of labour.

In the end we now have the following simplified system of 3 equations:

\[ H = d + e Q \]

\[ e = f(L, PVU, DYN, HS, DE_r, A, U) \]

\[ Q = f(I, PD, DE_r, COM, N_k, N_{rk}) \]

The endogenous variables are \(H\), \(e\) and \(Q\).

For all variables we now stand a fair chance of finding observed variables which are sufficiently close. This system therefore provides the basis for our subsequent empirical studies.

As a final question we should ask whether we have succeeded in our purpose of deriving a macro model.

To begin with, let us consider 6.3). It explains the average value of \(e\) for those shops in a given shopotype for which the conditions as specified in the exogenous variables are the same. The exogenous variables were introduced as averages for a region supplying customers to a given shopping centre, but they are equally meaningful (although perhaps less powerful) as averages for larger regions (states, provinces) and for a whole country. 6.3) can thus be interpreted as a function explaining the national average for \(e\) in a given shopotype. However, when we proceed to a time series study of national shopotype averages, the specification of 6.3) may be modified for that purpose.

For example, we may try to include some measure of any changes in the service
level (GSL), technical progress (θ) or changes in the division of labour between retailers and wholesalers/ producers (ψ).

Our macro version of 6.3) would then be:

6.4) \[ e_i = f(L, PVU, DYN, HS, DE_r, A, U, \ldots \ldots .) \]

Thus we are on our way to a macro model, even if its final specification is not ready, at least with respect to an explanation of sales efficiency, which was our main purpose.

Considering 5.9), the situation is different. 5.9) is based on a regional distribution of demand while what we will need is a distribution of aggregate demand between showtypes (intertype competition) and between shops in a given showtype (intratype competition). To achieve this we need a model to explain the number of shops in a given showtype \( N_i \) and the market share of showtype i in product classes \( r(\alpha_{ir}) \), as indicated in 5.3). Cross section studies on the basis of 5.9) within a showtype may have some value in a macro context, but it is largely a by product, and the main job in this respect is still to be done.

The basic linear cost structure 3.1) applies at any level, as long as we keep to a given showtype. Hence the total number of labour hours in showtype i is as follows:

6.5) \[ TH_i = N_i (d_i + e_i Q_i) \]

where:

- \( TH_i \) is total labour hours in showtype i
- \( Q_i \) is the average sales size in showtype i
- \( N_i \) is the number of shops in showtype i

An important implication of the linearity of the structure is that we do not need a distribution of sales sizes to arrive at the aggregate volume of labour. We do need such a distribution, however, if we want to explain the distribution of shopkeepers' income rather than only the average.

Before we proceed to a further development of the macro model we will first test the model given by 3.1), 6.3) and 5.9) on the basis of cross section studies with both individual shops and regional averages, for a few showtypes. Having tested and, possibly, modified or extended the model for e on that basis, we expect to have a better basis to proceed to a time series version of 6.3). We further note that the model is useful in its own right, as a tool to explain interregional differences in "productivity" and sales size for a given type of shop.


7700  List of Reprints, nos. 179-194; List of Reports, 1976
7701/M  "Triangular - Square - Pentagonal Numbers", by R.J. Stroeker.
7702/ES  "The Exact MSE-Efficiency of the General Ridge Estimator Relative to OLS", by R. Teekens and P.M.C. de Boer.
7703/ES  "A Note on the Estimation of the Parameters of a Multiplicative Allocation Model", by R. Teekens and R. Jansen.