Costs, margins and competition
Causes of structural change in retailing

Bart NOOTEBOOM *

The paper analyses causes of structural change in retailing in terms of productivity and profit. The analysis of productivity focuses on economics of scale, and briefly considers the effects of an extension of shopping hours. The analysis of profits considers the factors that determine average margin per type of trade. The results are used to explain trends of increasing scale, concentration and declining share of independents.

1. Introduction

In many countries we have witnessed the following phenomena in retailing: increasing scale of shops, decreasing density of shops, increasing concentration, and declining share of independents. For a descriptive account of these phenomena in several countries, see the contribution by Nooteboom, Thurik and Vollebregt in the present issue.

Underlying causes of these developments occur on both the demand and the supply side. Factors on the demand side are relatively well known. There are factors associated with a higher standard of living. One of these is consumers' greater mobility (car ownership), which allows for a greater distance and a lower density of shops. Another

factor is the increasing confidence in and familiarity with a widening range of goods, due to both a higher level of education and an increased volume and frequency of purchases, which allows for less service and private labels of multiples. This familiarity with products is enhanced by opportunities for mass advertising through the media. Scarcity of time [1], the rising proportion of women working outside the home, and more ample means and opportunities for recreation have yielded a preference for efficient shopping (self-service, one-stop-shopping), with large volumes of purchases per shopping trip, made possible by increased ownership of cars, refrigerators, freezers and other facilities for home storage.

Since those causes on the demand side are fairly well known, the present paper will not discuss them in any detail. Causes on the supply side are not all widely known, and some of them are controversial. Two controversial issues are price discrimination ('unfair' discounts from suppliers) and the effects of longer trading hours. The paper will briefly touch upon both issues. Research at the Research Institute for Small and Medium Sized Business in the Netherlands yields insight into causes associated with productivity, particularly economies of scale in store operation, and with the determination of retail margins. The present paper will focus on those less well known causes of structural change.

First the issue of productivity is discussed, in particular, differences in productivity between small and large shops of the same type.

[1] Recently, of course, there have been growing numbers of unemployed people, for whom time is not scarce, but who have less money to spend.

Implications for productivity of extended shopping hours are considered briefly. Next the determinants of profits are considered: How are profit margins affected by productivity, scale, type of trade, competition and the business cycle. In the final section the analysis of productivity and profit is used in an attempt to explain the phenomena of structural change in retailing in general and the food trade in particular.

2. Productivity

Unlike a factory, a retail unit does not physically produce a product, but provides a facility which consumers may decide to utilize. The causation involved is different. In retailing the initiative for the use of capacity lies on the side of the consumer, while in industry it lies on the side of the producer. Industry provides a utility of form, while retailing provides a utility of time and place. In view of those differences, one should not too readily and uncritically employ concepts and tools from studies of productivity in industry. The traditional concept of a production function in economic theory is of dubious relevance in retailing. It is true that the establishment of a shop attracts consumers and, in that sense,

can be said to generate sales with the means of capital (shopping space and inventory) and labour (for service or cashing). But one may also, and perhaps more relevantly, view retailing as adjusting capacity to a flow of demand. Taking the latter approach, Nooteboom (1980 : ch. 3, 1982) developed a theory of retailing costs based on the theory of queues. Given a certain pattern of consumer arrivals, and the time needed to serve customers (which depends on the volume of purchases per customer and the technology and quality of the service), how much labour is required to prevent waiting times on the part of customers that exceed the target, consistent with the service level chosen as part of the retailing mix?

Queuing theory was developed to answer this type of question. Of course, the intensity of customer arrivals varies during the day, the week and the year. That is where part-time labour comes in to adjust capacity to demand. The relationship between capacity (in terms of labour and shop space) and sales size per shop of a given type, i.e., with a given retailing mix, was determined and proved to be approximated by a simple straight line, as illustrated in fig. 1.

This linear cost curve has a threshold volume, i.e., a certain minimum which ob-
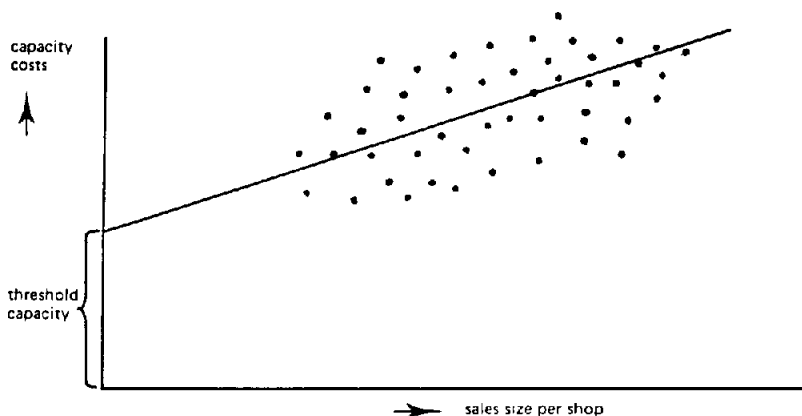

Fig. 1. Linear cost curve.

tains even if sales are zero. In the case of labour, the threshold is equal to the annual opening time of the shop, if we are considering costs on an annual basis. In the Netherlands the legal maximum of opening time is 52 hours per week, which per annum, allowing for holidays, yields about 2,500 hours. When we estimate the cost curve for the Netherlands for different types of trade, the threshold is consistently estimated at or just below this volume, which corroborates the underlying theory. In other countries with longer opening times, one expects and finds higher threshold values. The threshold applies if there is only a line of check-outs or a single service desk or counter. If there are more service units spread around the shop, the threshold must be multiplied accordingly. This also is confirmed in empirical tests of the theory (cf. Nooteboom, (1980, 1987)). Actual data are always scattered around the straight line, as illustrated in fig. 1. Partly, this scatter is explained by the theory, on the basis of differences in the use of part-time labour, quality of labour, costs of labour, volume of purchases per customer, service level, location and use of technology (type and extent of automation, for example) (cf. Nooteboom (1980, 1987) and Thurik (1984a)). To the extent that the shops considered are more similar in those respects, the scatter around the

curve is less. The model has been estimated and validated on data from the Netherlands, Germany, France, UK, USA, Canada and South Africa (cf. Nooteboom (1980, 1987) and Thurik (1984a)). For shop size, the threshold (for a single service unit) is estimated at about 50 m^2. [2]

The linear cost curve implies a considerable economy of scale at the lower end of shop size. For small shops the threshold costs weigh much more heavily than for large shops of the same type. Labour productivity, defined as sales per hour of labour, increases with shop size in the way illustrated in fig. 2. Studies of gross margins for shops of the same type show that percentage gross margin does not systematically depend on sales size. In other words, while for smaller shops percentage costs are higher (due to threshold costs), percentage margin is not. If percentage margin is not size dependent, the relation between gross profit and sales size is a straight line without intercept. The confrontation between gross profit and costs, to arrive at net profit as a function of size, thus is as illustrated in fig. 3. Due to threshold costs, net

---

[2] In better climates than in the Netherlands, with opportunities for queuing outside the cover of the shop, the threshold may be lower.
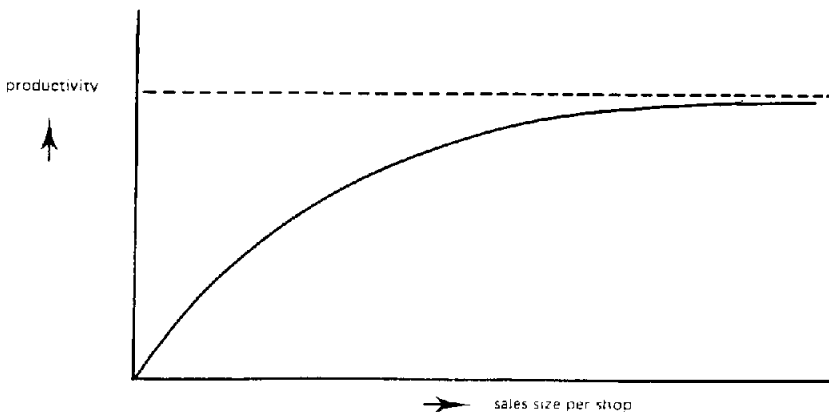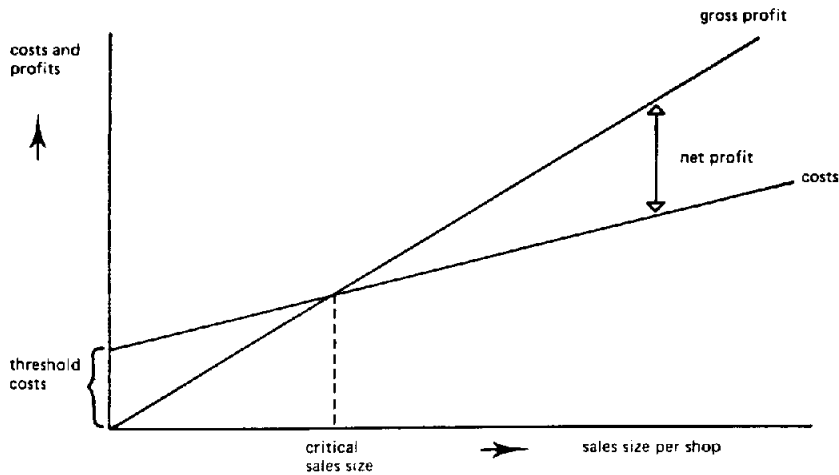

Fig. 2. Labour productivity.

Fig. 3. Net profits and size.

profit is positive only beyond some critical sales size.

3. Trading hours

Since the labour threshold equals opening time (for each department or service point in the shop), an extension of opening time will increase the threshold, thus shifting the cost curve upwards. If percentage margin remains the same, critical sales size will then increase, as illustrated in fig. 4. [3] To express it differently, an extension of opening time yields an increase of costs which in percentage terms is greater for smaller shops.

The conclusion is as follows: If the number of hours that shops are actually open increases, and if there is no increase or shift of sales in favour of small shops, there will be a further impulse toward an increase of scale and decline of small shops.

A more favorable effect for smaller shops would arise if an extension of shopping hours were to lead to a shift of sales to smaller shops. This may be the case for shops on the

neighborhood level, due to a shift to the market segment of purchases by women working outside the home and single workers outside regular business hours. On the other hand, the trend towards reduced working times may weaken that effect. The increase of sales required to offset increased threshold costs is a fixed amount, independent of sales size. [4] Thus the percentage increase of sales required is larger for smaller shops.

Of course, it is not logically necessarily so that a permitted extension of shopping hours will lead to actual lengthening of opening time. It might lead to a shift of opening time to different hours without an extension of total opening time. However, retailers tend to feel obliged by competitive pressure to extend hours when that is allowed. Evidence for this is found in the fact that average opening time is observed to be longer when allowed [5].

In view of these considerations, policy thinking in the Netherlands is inclined to go in the direction of maintenance of a maxi-

---

[3] Nooteboom (1983).

[4] If the formula of the cost curve is $c = d + eq$, where c = costs, d = threshold costs, q = sales size, and m = percentage gross margin, then the increase of sales required to offset an increase of threshold costs D equals $D/(m - e)$.

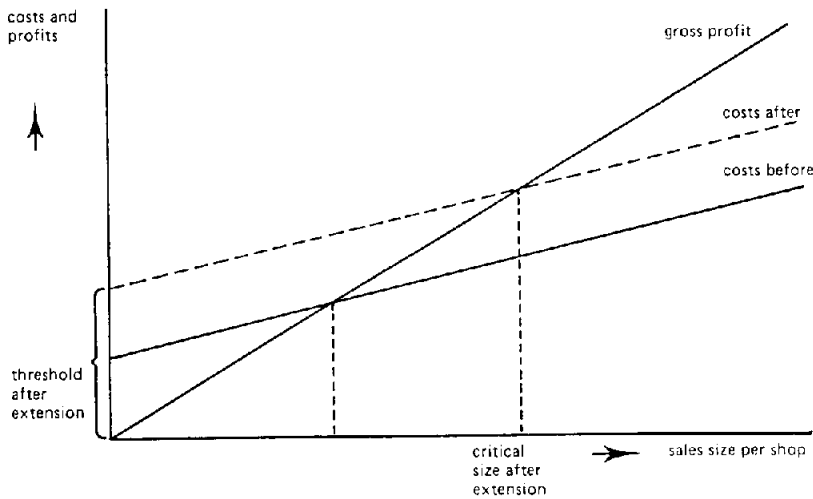[5] Cf. Nooteboom (1982) and Thurik (1984b).

Fig. 4. Extension of opening time.

mum opening time (at 52 hours per week), but with a certain amount of liberalization of the selection of opening times.

An important question is also whether for opening hours outside normal business hours labour can be obtained cheaply or only at a higher cost. This depends on regulations and agreements concerning wages and conditions of labour. On the whole, Western European countries have fewer opportunities for obtaining cheap, flexible, non-standard labour than the US, so that opportunities for changing opening hours without or with limited cost increases are less.

Independents are at an advantage here, to the extent that they can bring in informal labour of their spouses, children or other family members during nonstandard hours. (Of course, this is likely to detract from the liberty, leisure time and social contacts of the family, and may affect the homework and hence the education of the children.) Cultural developments in developed countries, with emancipation of women and children and a reduction of paternal authority, have reduced opportunities in this respect. Important exceptions are ethnic minorities where family

discipline and patriarchal rule are often still strong.

4. Profits

When percentage margins are compared between shops of different size but of the same type (same product/service mix), no systematic relation with shop size is found [6]. This implies that special discounts to larger shops, to the extent that they occur, tend to be passed on to the consumer. This is consistent with the conclusion of the report of a recent inquiry into the possibility of price discrimination in the U.K. [7] The report concludes that there are indeed special discounts to larger retailers, but that this does not matter since they are passed on to the consumer. But the point is, of course, that this offers the larger buyers an extra means to extend their market share, and yields a further increase of concentration, which further increases the buying power of the biggest companies.

[6] Cf. Nooteboom (1980: 67-68).
[7] Office of Fair Trading (1985).

A study of the factors that determine average percentage gross margin per type of trade is offered in Nooteboom (1980, ch. 8, 1985). The model developed in the study claims to explain both differences in average percentage margin between different types of trade, and the development in time per type of trade. The model was tested and estimated on Dutch and German data. In Nooteboom, Thurik and Vollebregt (1985) the model is applied also to other countries, in a comparison of margins between different European countries. Per type of trade, the model explains average percentage gross margin as a mark-up on average percentage operating costs excluding a reward for shopkeepers' labour. The height of the mark-up is explained by the following factors:

– Average shop size: the average percentage mark-up is inversely related to average sales size per shop. The rationale of this component of the model is that at a higher average sales size a lower percentage of sales is required to achieve a given amount of money as a reward for shopkeepers' labour. Estimation of the model implies an estimate of the average reward available for shopkeepers' labour. It turns out to be about equal, in the Netherlands, to the legal minimum wage for employees. This is the average reward; shops of above average size yield a reward which is proportionately higher and shops of below average size a reward which is proportionately less.
– Type of product and service level: a more varied range of products, slower-moving products and a higher service level require a higher profit mark-up to achieve a given return on capital (supply side argument), and justify a higher mark-up due to more service and a greater dependence of customers on the retailer (demand side argument).
– Stage in the life cycle of the shop type:

during the penetration phase of the retail formula (increasing market share) there is a Schumpeterian bonus for successful novelty; during the phase of saturation (constant market share) this bonus falls away; during the phase of decline (falling market share) the profit mark-up shrinks due to the competition of novel types of trade. The life cycle effect has been particularly prominent in the grocery trade, with its rise and decline of successive innovations: self service, supermarkets, discount stores, specialty supermarkets, soft discounters. [8]
– Business cycle: the level of the profit mark-up depends on the growth rate of consumer spending (in volume). The rationale is that in a contracting market, price competition is more intense, in the struggle for a larger share in a declining market to sustain sales volume, which leads to a lower average profit percentage. In Nooteboom and Thurik (1985) allowance was made for asymmetry in the business cycle effect: a difference in effect between recession and growth.

The mathematical formula of the margin model is given in the appendix.

5. The ousting of smallness

The economy of scale (higher costs for smaller shops), together with the way in which margins are determined, yields a mechanism by which smaller, independent shops tend to be systematically pressed out of the market. The basic driving force is illustrated in fig. 5.

Because costs have a threshold, while percentage margin does not depend on size, net profit is positive only beyond some critical sales size, as already indicated. The penalty
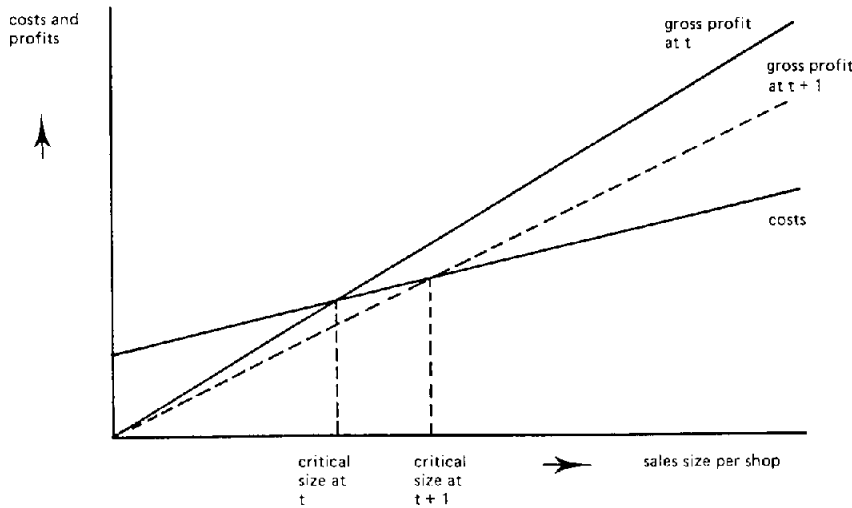
---

[8] See Nooteboom (1984).

Fig. 5. Increase of scale.

on small size is negative profits, while beyond critical size net profits increase in proportion to sales. This tends to push up average sales size per shop.

According to the margin model, this increase of scale exerts a downward pressure on the profit mark-up over costs, so that in fig. 5 the gross profit line rotates downward, whereby critical size increases, which reinforces the tendency of increasing scale. Thus the process of increasing scale feeds upon itself. This process is not explosive, however, but damped: the decrease of percentage margin becomes less and less as average sales size increases. Due to the effect of the type of product and service level, in the margin model, attempts are made to upgrade the product/ service mix with specialty or luxury products and better service.

This is the well-known phenomenon of 'trading up'. This strategy is risky, however, since the higher cost of retailing that results will tend to evoke new, cheaper types of retailing. This is the well-know concept of the 'wheel of retailing'. The creation of new forms of retailing is particularly rewarding in view of the life cycle effect: successful innovation

is rewarded by an extra profit mark-up. In a study of innovation in retailing (self service, supermarkets, discount stores), Nooteboom (1984) found that a few independents were the pioneering innovators, but the majority of independents were laggards, who adopted the innovation much later, to such an extent that they were still entering in the declining stage of the life cycle, after large business had started to move out. Similar evidence is found in Dawson (1981). According to the margin model, the profit mark-up is squeezed during the phase of decline. As a result, the late entry and the lagging behind of small independents also contributes to their decline.

According to the business cycle effect, a switch from growing to stagnant or declining consumer spending yields a downward pressure on the profit mark-up. This is what happened in the Netherlands, for example, when consumer income declined in the years 1980–1983. The resulting profit squeeze, yielding a further downward rotation of the profit line in fig. 5, with as a result a further increase of critical size, led to a acceleration of the decline of small shops.

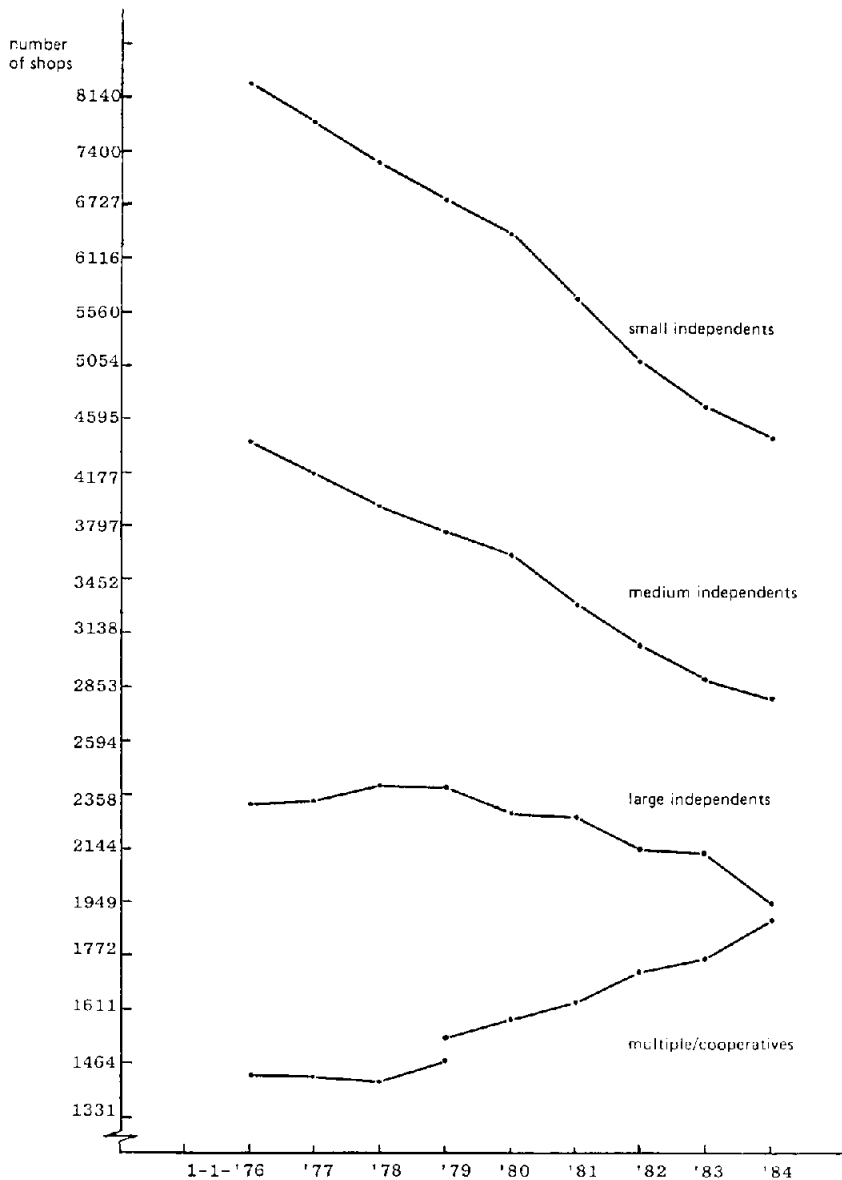This is illustrated in fig. 6, which shows,

Fig. 6. Number of shops in the grocery trade (logarithmic scale). One unit of scale along the vertical axis corresponds to a 10% change. Source: Nielsen Food Index.

for the Dutch grocery trade:

– a decline in the number of independent shops which is steepest for the smallest independents, and an increase in the number of shops of multiples.
– an acceleration of the decline of independents in 1980 (when consumer income

started to decline). In 1983–1984 the acceleration subsided, and the decline of shops appears to proceed at its previous pace.

It is in line with the margin model that the acceleration of the decline of smaller shops fades away. According to the model, the *level* of the mark-up depends on the *percentage*

change of spending. Thus the mark-up is reduced when a growth of spending switches to a decline, or when the decline rate increases, but it remains the same if the decline rate subsequently remains the same. This is what happend in 1980–1983. After 1980 there is no further downward pressure on margin, and hence no additional upward pressure on critical size. It will take some time for the upward shock of critical size in 1980 to work itself out, but after a while a certain recovery is to be expected. From 1984 onwards, consumer spending has again shown modest growth. According to the margin model, the transition from declining to slightly growing expenditure widens the profit mark-up, and may thus

reduce critical size. Thus the decline rate of independents may temporarily become less than it was prior to 1980.

Clearly, the decline of independents, combined with a rise of multiples, goes together with a decline of market share for independents and an increase for multiples. This is shown in fig. 7.

Thus it is shown that during the recession of the early eighties the market shares of multiples and independents have diverged dramatically. This is ascribed to a reinforcement of an existing trend, due to a further squeeze of profit margins as the result of a decline of consumer spending, which has led to an accelerated squeezing out of smaller
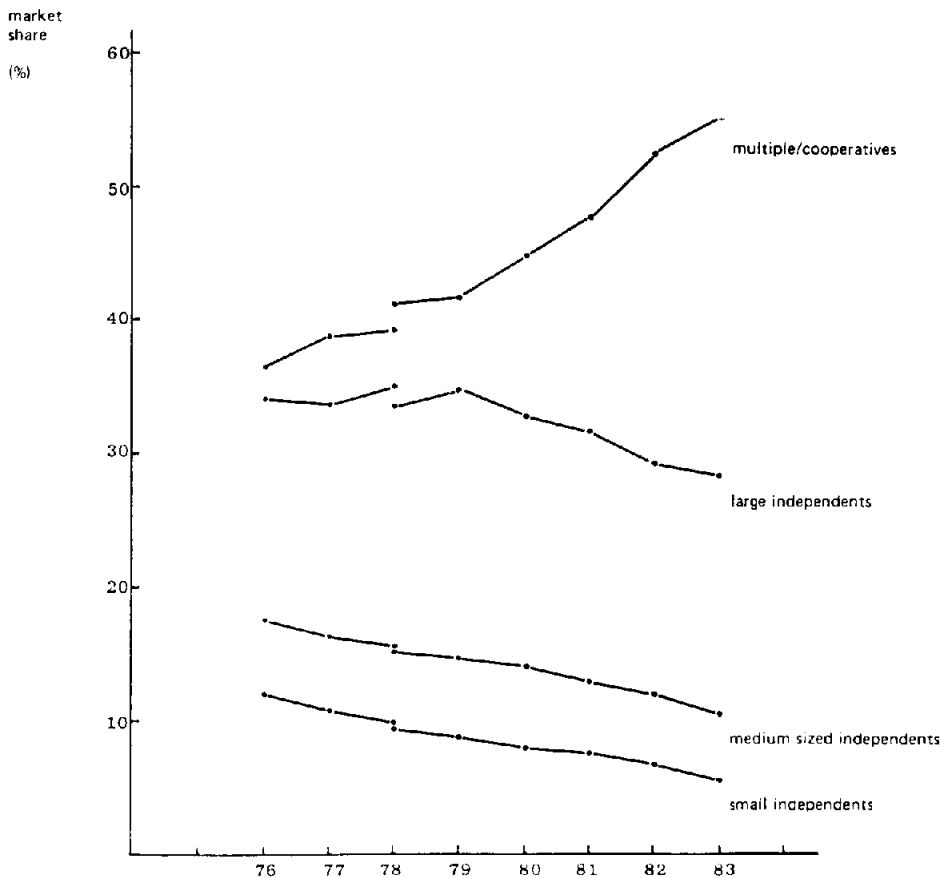

Fig. 7. Market shares in the grocery trade (linear scale). Source: Nielsen Food Index.

independents due to economy of scale. Perhaps this is not the full explanation. It it theoretically plausible – and there is empirical evidence – that there is price discrimination in the sense of excessive discounts to large buyers. It is also plausible that in years of stagnant or declining consumer expenditure the phenomenon is more pronounced, for two reasons:

- In a time of pressure on consumer prices, it is more tempting for large multiples to exert their buying power to the fullest.
- Manufacturers face excess production capacity, which makes them more vulnerable to pressure to deliver to special customers at low, marginal costs.

Thus price discrimination may also be subject to a business cycle effect, and this may have contributed to the phenomena of figs. 6 and 7.

Appendix

Formula of the margin model

In mathematical terms, the model is as follows

$$m_{it} = k_{it} + \underbrace{b_0\pi_t/Q}_{\begin{pmatrix}\text{scale}\\\text{effect}\end{pmatrix}} + \underbrace{a_2 e_{it}}_{\begin{pmatrix}\text{shoptype}\\\text{effect}\end{pmatrix}}$$

$$+ \underbrace{a_3 Wm_{it}}_{\begin{pmatrix}\text{lifecycle}\\\text{effect}\end{pmatrix}} + \underbrace{a_4 \overset{\shortmid\shortmid}{c}_{it}}_{\begin{pmatrix}\text{business}\\\text{cycle}\\\text{effect}\end{pmatrix}} + U_{it},$$

$$\overline{\hspace{4cm}\text{mark-up effect}\hspace{4cm}}$$

where

i	=	index of shop type,
t	=	index of the year,
m	=	average percentage gross margin,
k	=	average percentage operating costs (excluding a reward for shopkeepers' labour),
b_0	=	average reward for shopkeepers' labour in some base year t_0 (to be estimated),

π	=	cost of living index with respect to base year t_0,
Q	=	average sales size per shop (at current prices),
a_2, a_3, a_4	=	coefficients estimated (all are positive),
e	=	income elasticity of the product/service package offered, as a measure of its level of luxury/specialty,
Wm	=	change of market share of shoptype,
$\overset{\shortmid\shortmid}{c}$	=	percentage change of consumer spending (in volume), ad
U	=	stochastic disturbance term.

For the Netherlands the estimates of the coefficients are as follows:

$b_0 = 7\frac{1}{2}$ thousand guilders (base year 1970), which was about equal to the legal minimum wage of employees,

$a_2 = 4.0$,

$a_3 = 0.5$, and

$a_4 = 0.11$.

References

Dawson, J.A., 1981. Innovation adoption in food retailing – The example of self service methods. Service Industries Review 1, 22-35.

Nooteboom, B., 1980. Retailing: Applied analysis in the theory of the firm. Amsterdam: J.C. Gieben.

Nooteboom, B., 1982. A new theory of retailing costs. European Economic Review 17, 163-186.

Nooteboom, B., 1983. Trading hours and economy of scale in retailing. European Small Business Journal 1, 57-62.

Nooteboom, B., 1984. Innovation, life cycle and the role of independents: Cases from retailing. International Small Business Journal 3, 21-33.

Nooteboom, B., 1985. A mark-up model of retail margins. Applied Economics 17, 647-667.

Nooteboom, B., 1987. Threshold costs in service industries, The Service Industries Journal 1, no. 1, 65-76.

Nooteboom, B. and A.R. Thurik, 1985. Retail margins during recession and growth. Economics Letters 17, 281-284.

Nooteboom, B., A.R. Thurik and J.A.C. Vollebregt, 1985. Do retail margins differ among European countries? Proceedings of the CESCOM-IRF Conference on Distribution in Italy, April 19, 1985.

Office of Fair Trading: Competition and Retailing, June 1985, U.K.

Thurik, A.R., 1984a. Quantitative analysis of retail productivity. Delft: Meinema.

Thurik, A.R., 1984b. Labour productivity, economies of scale and opening time in large retail establishments. The Service Industries Journal 4, 19-29.