Percentage Retail Mark-Ups

THOMAS VON UNGERN-STERNBERG*

1. INTRODUCTION

Introductory courses on industrial organisation frequently teach the «double marginalization» problem. The basic message is that a chain of monopolies charges a final price that is higher than the one which would maximize joint profits. The simplest example, and the one on which we shall concentrate from now on, is when a good passes from a (monopoly) producer to a (monopoly) retailer who then sells it to the final customer. In the standard textbook models the producer first adds a given margin (or mark-up) to his marginal costs, the retailer does the same to his wholesale price, and the resulting (Nash or subgame perfect) equilibrium outcome is a consumer price that is inefficiently high even when only the supply side of the market is taken into account.

This inefficiency is then interpreted either as a motive for vertical integration or as a reason for using franchise fees rather than (or in combination with) linear pricing.[1] The reasons evoked for studying franchise fees and/or linear pricing is, that they are both simple and the mechanisms most frequently observed in the real world. I agree with the first point and consider that much could be gained if one concentrated more on studying simple rather than «optimal» incentive mechanisms. The second claim is not quite correct. What one does observe in the real world is that retailers usually charge a mark-up that is a given *percentage* of their wholesale prices. This mark-up has to cover both their variable and their fixed costs (and perhaps yield a little profit as well). A percentage mark-up is not, of course, the same thing as an absolute mark-up, and has quite different incentive effects. The purpose of this paper is to explore these differences and explain why one observes percentage mark-ups so frequently in practice.

There is substantial evidence, that the empirical claim I make is correct: SCHERER and ROSS write: «In the retail trades, a conventional pricing rule is to seek some standard percentage margin – for example 40% – of price less cost over price. Knowing the wholesale price W of an item, one finds the retail price by calculating W/(1–.4). The 40%

* University of Lausanne
I would like to thank Andreas Irmen and Damien Neven for very helpful comments and suggestions. I am also deeply indebted to an anoymous referee from this journal for exceptionally detailed and helpful comments. I accept responsibility for any remaining errors.

1. C.f. SPENGLER (1950).

margin in this case must cover all selling and overhead expenses.»[2] I have telephoned with the Migros, Switzerland's biggest retail chain, with an annual turnover of 12.5 billion Swiss francs, and they confirm, that they use exactly the pricing policy described by SCHERER and ROSS. The percentage margins they work with are not, of course, the same over all product lines. (For example they are higher for frozen food than for tinned food because of the higher variable and infrastructure cost.) The essential point is, however, that any supplier working in a given segment of the market knows that his wholesale price will be multiplied by a given coefficient to determine the final consumer price. In other industries such as the film industry (on the continent but not in the UK) it is the producers who claim a given percentage of the final retail price. (This has similar incentive effects but in the other direction.) Finally it is well known, that the world's two greatest auction house (Christies and Sothebys) operate with fixed percentage mark-up (for objects within a given value range). It thus seems empirically relevant to examine the essential differences between the text-book linear pricing policies and the percentage mark-up policies one observes in the real world.

The idea of studying the effects of percentage retail mark-ups occurred to me as a result of the following personal experience: My wife produces glass objects which she wanted to sell through an art gallery in Zurich. The owner of the gallery told her, that he pursued the following simple pricing policy: The retail price he would ask was the «producer» price my wife charges multiplied by a factor of two. It was easy to see that this left my wife with essentially the following trade-off: Either she charged a producer price which left her a reasonable margin; the final consumer price would then be so high, that the sales volume would tend to zero. Alternatively, she could ask a very low producer price which would leave her with practically no margin, but at least the sales volume would then be positive. All attempts to get the gallery to change his pricing policy were of no avail. In the end no transaction took place.

This simple example should suffice to illustrate one of the main effects of percentage mark-up policies. *The retailer who commits to such a policy is squeezing the producer's margin.* He sets the producer a strong incentive to charge low wholesale prices because this will translate into low retail prices (and higher sales volumes).

The most obvious way to model this interaction is in the form of a two stage game. *In the first stage the retailer fixes his percentage mark-up and in the second stage the producer then sets his wholesale price.* It has been suggested to me several times, that this is not the «natural» way to model the sequence. There seems to be a wide spread belief that «the natural assumption in the timing structure is that the producer chooses first and the retailer second, since this is the chronology of the product's movement towards the consumer».[3] I cannot agree with this point of view. In the real world the

2. SCHERER and ROSS (1990), 262. In their formulation, the margin is calculated as a fraction of the retail price. In our model, the margin will be calculated as a fraction of the wholesale price. The two approaches are of course equivalent.

3. Quote from an anonymous referee report from the IJIO.

interactions between a producer and a retailer are practically invariably of a repeated nature. If one attempts to model the outcome of such interactions by using a static two stage game, one presumably wishes to give the «first mover advantage» to the player with the greater bargaining power. Surely, the way profits are shared depend more on the players' relative «bargaining power», than on the question of whether they are up-stream or down-stream firms. One of the main preoccupations of competition authorities in many industrialised countries is the growing purchasing power retailers can exert vis-à-vis the producers. In those situations the «natural» assumption is that the retailer is the leader and the producer the follower.

The effect of working with percentage mark-ups can be understood from a simple two stage model where a producer sells a good to a retailer who then sells it to the final consumer. In the first stage the retailer fixes his (percentage or absolute) mark-up, in the second stage the producer sets his wholesale price: By fixing a high percentage mark-up (say 50%) in the first stage, the retailer essentially signals to the producer that every time he increases his price by 1$ the retail price will increase by 1.50 $. The reduction in sales volume this implies is, of course, a strong incentive for the producer to keep his wholesale price low. A first obvious conclusion is thus that percentage mark-ups set by the retailers in the first stage allow to squeeze the producers' mark-up in the second stage. Comparing the results of this game with the standard linear pricing version we find that:

a) producer mark-ups are lower
b) consumer prices are lower
c) retailer profits are higher, and
d) aggregate (producer + retailer) profits are higher.

Whether the absolute value of the retailer's mark-up is higher or lower is ambiguous. The same is true for producer profits.

Points b) and d) are the most interesting ones. They imply that percentage retail mark-ups partially solve the double marginalization problem. The retail price does not reach but moves in the direction of the vertically integrated monopoly price.

It should be noted, that these results do *not* depend on the two stage sequence. They also hold, if the producer and retailer move *simultaneously* rather than sequentially.

Let us now turn to the next part of the analysis. The fact that percentage retail mark-ups increase aggregate profits obviously implies that *both* retailers and producers could gain from such a strategy if only the profits were divided differently. Suppose we change the game so that there is competition among retailers. This forces their profits to be zero. It will then be the *producer* who benefits from the higher aggregate profits. If aggregate profits are higher with percentage retail mark-ups, the producer will prefer to work with a retailer using such a strategy. This is the second main result of the paper: If there is strong retailer competition it may be the producer who *chooses to* work only with retailers charging percentage mark-ups. It is not so much the retailers who squeeze the producer by working with percentage mark-ups, but the producer who may accept only to work with such retailers. The reason for this may seem somewhat paradoxical. A retailer working with absolute mark-ups gives the producer insufficient reasons to keep his own wholesale prices low. This exacerbates the problem of double marginalisation. The producer may wish to select a retailer whose mark-up policy gives him an incentive to charge a lower mark-up than he would under linear pricing. Since this may have the effect of letting the final consumer price approach the one that maximises joint profits, the result may not be quite so surprising after all.

All the results described so far were based on the assumption that there was no possibility to charge a franchise fee. As a result of restricting the strategy space in this way the outcomes achieved were never first best (i.e. joint profit maximising). In view of the fact that franchise fees are observed only quite rarely in the real world, studying such restrictive models seems to me to be a worthwhile exercise. Other economists consider that one should restrict oneself to «optimal contracts». The purpose of the last section is therefore to point out that the preceding analysis is of direct consequence for the design of optimal contracts. In the standard absolute mark-up models the simplest way to achieve the first best contract is for the party moving in the first stage is to sell to the second stage player at marginal cost, and to extract all the profit by charging a sufficiently high franchise fee. It is widely recognised that this approach places all the risk with the second player, and may create moral hazard problems for the first player. If the player in the first stage charges a percentage mark-up these problems are partially solved. Since the percentage mark-up makes the demand curve facing the second stage player more elastic (as compared to an absolute mark-up), the first stage player can now work with a positive mark-up, and the final price will still not exceed the joint profit maximising price. The contract which maximizes joint profits still contains a franchise fee, but it is also characterised by the fact, that both the producer's and the retailer's per unit mark-ups are positive. The risk is spread more evenly and the moral hazard problems for the first stage player are reduced. To put the same point differently: With percentage mark-ups, one can always design a contract which achieves joint profit maximisation, but which nevertheless has both parties' profits increase with total sales volume. This is attractive both for incentive reasons and if the second stage player is risk averse.

The rest of this paper is organized as follows: Section 2 sets out the assumptions of the model and solves it for the case where there is no retailer competition. Section 3 introduces the effects of retailer competition. Section 4 discusses problems of risk sharing and the design of first best (joint profit maximising) contracts. Section 5 ends with some concluding remarks.

2. THE MODEL

2.1. Assumptions

Demand

The demand function is assumed to be of the constant elasticity type:

$$x = Ap^{-\varepsilon} \qquad \varepsilon > 2.$$

[Working with a constant elasticity demand simplifies the analysis even for the standard linear mark-up model (c.f. TIROLE 1988, 176). It is the only formulation which allows us to obtain explicit equations when working with percentage mark-ups. One requires an elasticity of demand greater than two in order to obtain interior solutions.]

Without loss of generality we shall normalise the units of output such that $A = 1$, i.e. the demand function is

$$x = p^{-\varepsilon} \qquad \varepsilon > 2. \tag{1}$$

Supply

Production costs are of the simple constant marginal cost type, i.e.

$$CT_P(x) = c\, x. \tag{2}$$

(Introducing fixed production costs would not modify the analysis in any way).

Retailing costs are also assumed to be of the constant marginal cost type, i.e.

$$CT_R\,(x) = F + k\, x. \tag{3}$$

(The retailer fixed costs will play and important role when analysing competition among retailers in section 3).

We assume throughout that the *producer* (moving in the second stage of the game) charges a given absolute mark-up over his production costs. This is done only to keep the notation as conventional as possible. The reader can easily convince himself that none of the results would change in any way if the producer worked with percentage mark-ups over marginal costs. (Any given absolute mark-up automatically translates into a percentage and vice-versa.)

2.2. *Simultaneous Games*

2.2.1 *Absolute Mark-Ups*[4]

Let us start off with the model, where the producer and the retailer move simultaneously. Denote by m and n the absolute mark-ups charged by the producer and the retailer respectively. The retail price will then equal

$$p^{sim} = c + k + m + n. \tag{4}$$

Producer profits are

$$\Pi_P\,sim = m(c + k + m + n)^{-\varepsilon}. \tag{5}$$

Retailer profits are

$$\Pi_R{}^{sim} = n(c + k + m + n)^{-\varepsilon} - F. \tag{6}$$

The first order conditions yield the reaction functions:

$$m^{sim} = (c + k + n)/(\varepsilon - 1) \tag{7}$$

and

$$n^{sim} = (c + k + m)/(\varepsilon - 1). \tag{7'}$$

The symmetric equilibrium is reached when

$$n^{sim} = m^{sim} = (c + k)/(\varepsilon - 2). \tag{8}$$

The equilibrium price equals

$$p^{sim} = (c + k)/(1 - 2/\varepsilon). \tag{9}$$

This price obviously exceeds the price p* that would maximize joint profits

$$p^* = (c + k)/(1 - 1/\varepsilon). \tag{10}$$

4. The analysis of sections 2.2.1 and 2.3.1 is the same as the analysis in TIROLE (1988), 176.

2.2.2 *Percentage retail mark-ups*

Let us now compare this solution with the case where the retailer's policy consists of choosing an optimal *percentage* mark-up *r* over his wholesale price.

By definition the wholesale price equals:

$$p_w = (c + m). \tag{11}$$

The retail price thus equals

$$p = (1 + r)(c + m). \tag{12}$$

One notes that the retailer will have to set his percentage mark-up r so as to cover both his variable costs k and his fixed costs F (and sometimes leave him with some positive profits).

The producer's profit maximisation problem is now to

$$\max_{m} \Pi_P{}^{\text{sim\%}} = m\,[(1 + r)(c + m)]^{-\varepsilon}. \tag{13}$$

The retailer's profit maximisation problem is to

$$\max_{r} \Pi_R{}^{\text{sim\%}} = [r(c + m) - k]\,[(1 + r)(c + m)]^{-\varepsilon} - F. \tag{14}$$

The first order conditions yield:

$$m^{\text{sim\%}} = c/(\varepsilon - 1) \tag{15}$$

and

$$r^{\text{sim\%}} = (p_w + \varepsilon k)/p_w(\varepsilon - 1) \tag{16}$$

or

$$r^{\text{sim\%}} = [(c + m) + \varepsilon k]/\,[(c + m)((\varepsilon - 1)]. \tag{16'}$$

Substituting (15) into (16') one notes that:

$$r^{\text{sim\%}} = 1/(\varepsilon - 1) + k/c. \tag{17}$$

In absolute terms the retailer's mark-up is:

$$n^{sim\%} = r^{sim\%}\,(c + m) - k = \varepsilon c/(\varepsilon - 1)^2 + k/(\varepsilon - 1). \tag{18}$$

The equilibrium price is thus equal to:

$$p^{sim\%} = c/[1 - 1/\varepsilon]^2 + k/\,[1 - 1/\varepsilon]. \tag{19}$$

2.2.3 Comparison

Comparing the results obtained with absolute and percentage retail mark-ups, we find that the percentage mark-up strategy leads to

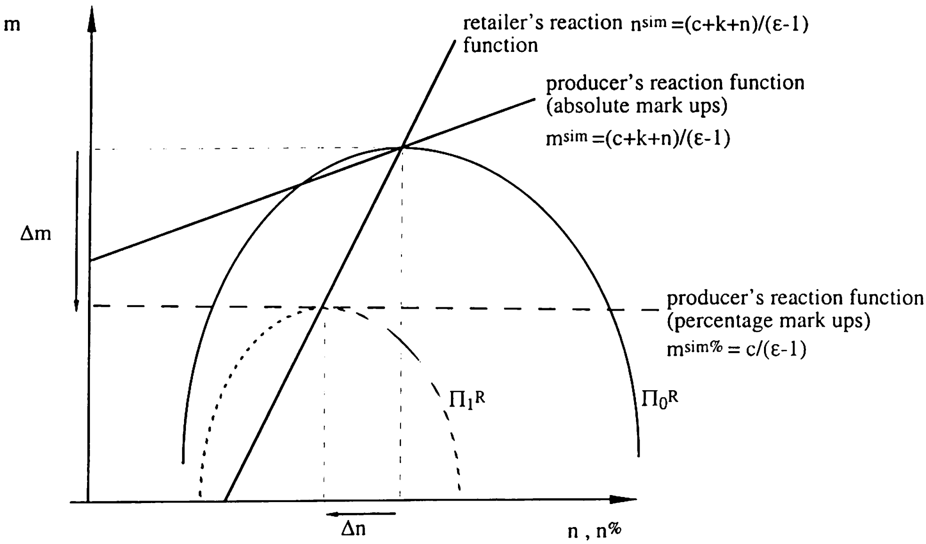
- a lower consumer price [c.f. (9) and (19)]
- a consumer price higher than the joint profit maximising price [c.f. (10) and (19)]
- a lower producer mark-up [c.f. (8) and (15)]
- a lower retail mark-up [c.f. (8) and (18)]
- higher aggregate profits and higher retailer profits.

The reason why percentage retail mark-ups lead to lower consumer prices and higher aggregate profits can be explained as follows: When the retailer charges a given absolute mark-up, this has the effect of shifting the demand curve the producer faces down by an *absolute* amount equal to this mark-up. The elasticity of demand decreases. When the retailer charges a given percentage mark-up, this has the effect of letting the demand curve «shift» down by a given *proportional* amount. The elasticity of demand remains unchanged. Since the producer faces a more elastic demand curve when the retailer charges a percentage mark-up (for any given absolute level of the retailer's mark-up), the producer reaction function is «rotated» downward, and he charges a lower wholesale price. Since the retailer's optimal mark-up is also lower, both mark-ups shift in the direction required to increase aggregate profits. One can easily check that the retailer's profits unambiguously increase. The only ambiguity in this model is whether the producer's profits rise or fall. On the one hand he now works with a much lower mark-up. On the other hand he obtains a higher sales volume. It is unclear which of these two effects is more important.

The situation is diagramatically depicted in *figure 1*[5] (Π_0^R and Π_1^R are retailer iso-profit curves).

5. Since we work with constant elasticity demand curves the absolute mark-ups are strategic complements, and the reaction functions in *figure 1* have positive slopes. With linear demand curves the reaction functions would have negative slopes (strategic substitutes).

Figure 1



2.3. *Two Stage Games: The Retailer Moves First*

Let us now consider the situation, where the producer must consider the retailer mark-up as given when fixing his own wholesale price. This is a reasonable assumption, when one sells through Sothebys or Christies, who always charge the same percentage mark-ups. By analogy it is also a reasonable assumption for markets where retailers sell a wide variety of products, and operate by adding a simple (percentage or absolute) mark-up to all their wholesale prices for a given type of good. The producer may then consider this mark-up as given when fixing his own wholesale price (unless, of course, he offers exceptionally high quantity discounts which may lead to special offers etc.).

2.3.1 Absolute Retailer Mark-Ups

The producer in the second stage of the game takes the retailer's mark-up as given when setting his own mark-up. His reaction function is thus identical to the one calculated in the simultaneous game [c.f. eq.(7)].

Substituting (7) into the retailer's profit function (6) we find that he will fix n so as to

$$\max_{n} \Pi_R{}^{seq} = n[(c + k + n)\,\varepsilon/(\varepsilon - 1)]^{-\varepsilon} - F. \tag{20}$$

The retailer's optimal mark-up is

$$n^{seq} = (c + k)/(\varepsilon - 1). \tag{21}$$

Substituting (21) back into the producer's reaction function (7) we find that the producer's mark- up is equal to

$$m^{seq} = (c + k)\ \varepsilon/(\varepsilon - 1)^2. \tag{22}$$

The price the consumers pay is

$$p^{seq} = (c + k)/(1 - 1/\varepsilon)^2. \tag{23}$$

This is unambiguously lower than the corresponding price in the simultaneous move game [c.f. (10)]. This result is well known, c.f. TIROLE (1988), 176].

2.3.2 *Percentage Retail Mark-Ups*

When the retailer in the first stage charges a given percentage mark-up, the producer's reaction function is again the same as in the corresponding simultaneous move game i.e.

$$m^{seq\%} = m^{sim\%} = c/(\varepsilon - 1) \tag{15}$$

Since the RHS of (15) is independent of the retail mark-up r, the retailer's optimization problem also is the same as in the simultaneous move game, i.e.

$$n^{seq\%} = r^{seq\%}\ (c + m) - k = \varepsilon c/(\varepsilon - 1)^2 + k/(\varepsilon - 1) \tag{18}$$

and the consumer price is once again equal to

$$p^{seq\%} = c/[1 - 1/\varepsilon]^2 + k/\ [1 - 1/\varepsilon]. \tag{19}$$

2.3.3 *Comparison*

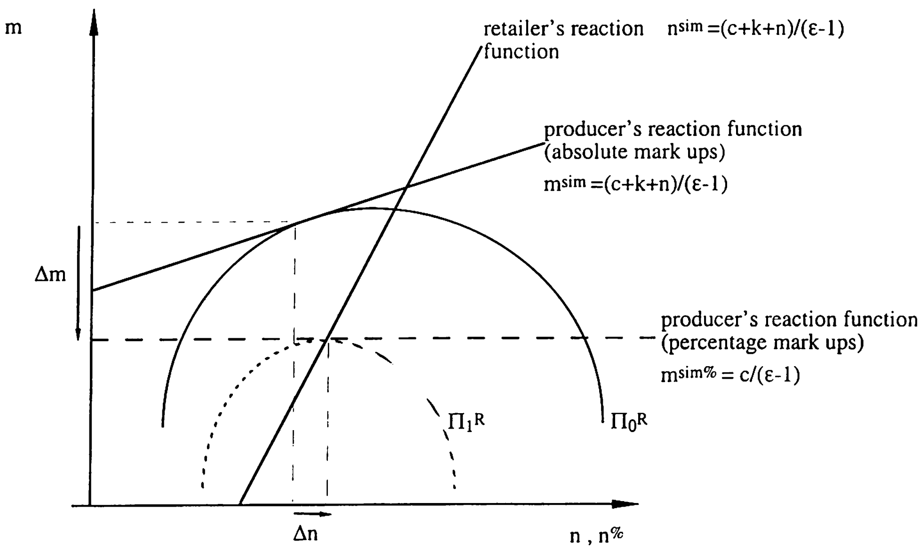
Comparing the results obtained with absolute and percentage retail mark-ups, we find that the percentage mark-up strategy leads to

- a lower consumer price (unless we have k = 0) [c.f. (19) and (23)]
- a lower producer mark-up [c.f. (15) and (22)]

- a higher retail mark-up [c.f. (18) and (21)]
- higher aggregate profits and higher retailer profits.

Once again the effect on producer profits is ambiguous. He sells a larger quantity at a lower mark-up. The situation is graphically depicted in *figure 2.*

Figure 2



The interpretation of the results is similar to the simultaneous game. There is, however, one difference. In the sequential game the percentage mark-up strategy leads to a *higher* retail mark-up, in the simultaneous case it leads to a *lower* retail mark-up. This difference can be explained as follows: The size of the retailer's mark-up is determined by the trade-off between his wish to keep the final price low (close to the profit maximising price) and his desire to squeeze the producer's mark-up. In the sequential setting the retailer has more scope for squeezing the producer's mark-up. In the absolute mark-up game he reacts by setting a lower mark-up than in the simultaneous game. With percentage mark-ups, the retailer can squeeze the producer's mark-up quiete efficiently even in the simultaneous game. He thus has less incentive to further lower his mark-up in the sequential game. The difference can be seen by comparing *figures 1* and *2.*

2.4. *Combining absolute and percentage mark-ups*

The analysis so far was based on the assumption that the retailer could choose to work with either an absolute or a percentage mark-up. Suppose now he could also decide to work with some linear combination of the two. The purpose of this section is to show that the retailer would still decide to work with percentage mark-ups only. (Franchise fees will be introduced in section 4.)

Assume that the retailer's policy consists of charging some percentage mark-up r combined with some absolute mark-up m. The producer's problem is then to choose his mark-up m so as to:

$$\max_{m} \Pi_P = m\,[(1+r)(c+m)+n]^{-\varepsilon}. \tag{24}$$

The first order condition is

$$m = [c/(\varepsilon-1)] + [n/(1+r)(\varepsilon-1)]. \tag{25}$$

m is clearly strictly increasing in n and strictly decreasing in r. The retailer's optimal choice is thus to set n equal to zero (negative values are excluded) and to make all his profits by charging a sufficiently high percentage mark-up. This result may also help to explain why the retailers do not add their own marginal cost k to their wholesale price before adding their percentage mark-up. Doing so would mean that they would have to work with a lower value of r to obtain a given absolute mark-up, and this in turn would decrease the pressure on the producers to keep their wholesale prices down.

2.5. *Two Stage Games: The Producer Moves First*

When the producer moves in the first stage of the game, the retailer no longer has the possibility to commit to any kind of strategy. In this situation it obviously makes no difference whether the retailer works with a percentage or an absolute mark-up. Any percentage mark-up translates into a given absolute mark-up. It is intuitively clear that the difference between percentage and absolute mark-ups is relevant only as long as *the producer cannot commit to a wholesale price before the retailer sets his mark-up.*

There is thus no point in writing down the equations for the situation where the producer moves first. The results will be the same, no matter whether the retailer works with absolute or percentage mark-ups.

3. RETAILER COMPETITION

The main result of the analysis so far was to show, that both aggregate profits and retailer profits increase, when the retailer adopts a percentage mark-up strategy. This result would seem to imply that the producer will try to prevent the retailers from committing to percentage mark-ups. Why then are they observed so regularly in the real world? To understand this, one must introduce the possibility of competition among retailers. This competition has the effect of limiting the absolute magnitude of the mark-ups they can charge. It may then well be the producer who benefits from the retailer's adopting a percentage mark-up strategy. Competition among the retailers may exclude from the market those retailers who try to operate with an absolute mark-up strategy.

There are several ways of modelling retailer competition. One possibility would be to have the producer supply several retailers in the same geographic area. Price competition among the retailers would then drive down their mark-ups. An alternative approach would be to have the producer work with only one retailer. Retailer competition would then take the form of letting the producer choose to work with that retailer who grants him the most favourable conditions. We shall adopt this second approach both because it is technically simpler and because it permits to highlight more clearly the main interactions we are interested in.

The model we examine consists of two stages: In the first stage two or more potential retailers fix the percentage or absolute mark-ups they would charge if they were granted the exclusivity. In the second stage the producer chooses the retailer with whom he will work and his wholesale price. As in the previous section, no lump-sum payments between the producer and the retailer are allowed. The main purpose of introducing competition among retailers in the first stage is that this will drive their profits down to zero. Their mark-up over wholesale prices will be just sufficient to cover their fixed and variable selling costs. The results of introducing this modification are as follows: When retailer fixed costs are «high» relatively to their variable costs, percentage mark-ups lead to higher aggregate profits than absolute mark-ups. The producer will then choose to work with a retailer charging percentage mark-ups. The opposite is true if retailer variable costs are relatively high.

3.1. A Solving the Model

One can of course try to explicitly solve this model in the standard manner by backward induction, and then compare the producer profits under absolute and percentage mark-ups. Unfortunately the expressions one obtains when proceeding in this manner are not particularly easy to interpret. A more indirect approach is indicated:

Denote by $x_{\%}$ the final output under percentage pricing and by x_a the output under absolute mark-up pricing. Similarly denote by

$$\Delta_{pa} = p_a - p_{wa} \tag{26}$$

and by

$$\Delta_{p\%} = r(p_{w\%}) = p_{\%} - p_{w\%} \tag{27}$$

the absolute values of the mark-up over wholesale prices charged by the retailer. The zero profit condition for the retailers then implies:

$$\Delta_{pa} = k + F/x_a \tag{28}$$

and

$$\Delta_{p\%} = k + F/x_{\%} \tag{29}$$

i.e. the retailer mark-up over wholesale prices must in both cases be just sufficient to cover their variable plus their per-unit fixed costs.

From section 2.3.1 we know that the producer mark-ups in the second stage of the game will respectively equal

$$m_a = (c + k + F/x_a)/(\varepsilon - 1) \tag{30}$$

and

$$m_{\%} = c/(\varepsilon - 1) \tag{31}$$

when the retailer charges absolute or percentage mark-ups.

Comparing (30) and (31) we see that the producer's mark-up will always be lower if the retailer works with a percentage mark-up strategy. This lower mark-up will translate into higher output and thus lower wholesale prices [c.f. (28) and (29)].

To study whether the producer prefers absolute or percentage mark-ups we thus have to examine under what circumstances higher total output (and lower consumer prices) lead to higher aggregate profits. (Since retailer profits are equal to zero, higher aggregate profits automatically imply higher producer profits.)

a) Let us start with the polar case where both F and k are equal to zero, i.e. the retailer has neither fixed nor variable costs. Both the percentage and the absolute mark-up of the retailers will equal zero and the producer will directly charge the joint profit maximising price

$$p_a = p_{\%} = c/(1 - 1/\varepsilon). \tag{32}$$

b) With $F = 0$ and $k > 0$ the price with absolute retail mark-ups is again equal to the joint profit maximising price

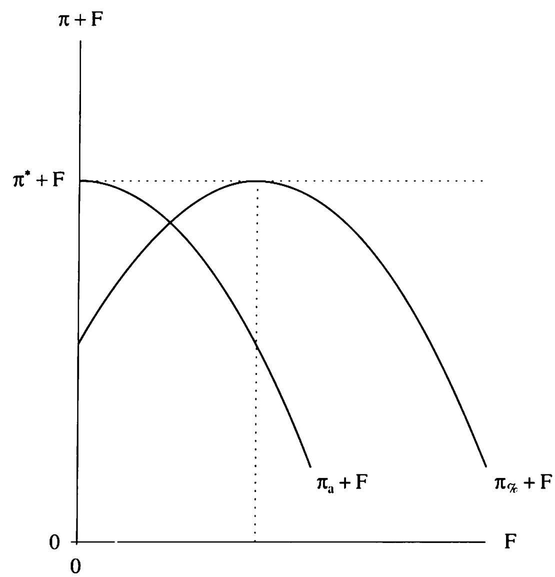
$$p_a = (c + k)/(1 - 1/\varepsilon). \tag{33}$$

The price with percentage retail mark-ups would be equal to

$$P_{\%} = c/(1 - 1/\varepsilon) + k \tag{34}$$

which is unambiguously *lower* than the joint profit maximising price. This is due to the fact that the percentage mark-up strategy gives the producer an incentive to base his own mark-up only on his own marginal production costs [cf. eq (15)], and the retailer can add only his own marginal cost k rather than $k/(1 - 1/\varepsilon)$ to his wholesale price.

Figure 3



c) Starting from a situation with $k > 0$ and $F = 0$ let us now increase the retailer's fixed costs to $F > 0$. With absolute retail mark-ups, p_a will now increase above the profit maximising level, i.e. the producer's profits will start declining (c.f. *fig. 4*). Similarly the price with percentage retail mark-ups will also increase. Since it is too low when $F = 0$ [c.f. eq. (34)] it will first tend toward the profit maximising price. It will reach this level when

$$F/x^* = k/(\varepsilon - 1) \tag{35}$$

i.e. when the per unit fixed costs are just sufficient to cover the insufficient mark-up charged by the producer. (x^* is the quantity that maximizes joint profits.) From then on $\Pi_{\%}$ also declines. The situation is graphically depicted in *figure 3.* Obviously for some level of fixed costs smaller than $F = x^* k/(\varepsilon - 1)$ the two profit functions intersect and percentage mark-ups yield higher aggregate profits from then on. It is unfortunately not possible to explicitly compute this critical level of fixed costs F as a function of k and. Roughly speaking one can merely say that percentage mark-ups are preferable to fixed mark-ups when fixed retail costs are «high» relative to variable retail costs and vice versa.

4. FRANCHISE FEES

In the analysis so far the retailer's and producer's contracts were limited to setting a simple absolute or percentage mark-up. As a result of this restriction the joint profit maximising outcome could not usually be attained. It is well known from the literature that with lump sum transfers (franchise fees) this aim can be achieved. In the two stage model of section 2.3 the retailer would set a mark-up equal to his own variable selling cost k, and extract all the producer's surplus by charging an appropriate fixed fee. Such a scheme has disadvantages as soon as one introduces risk aversion or moral hazard into the model. One party bears all the risk, and the other party faces moral hazard problems.

If the retailer has the possibility of charging a percentage mark-up the situation changes. The percentage mark-up increases the elasticity of the demand curve the producer faces (relatively to the situation with absolute mark-ups) and this pushes wholesale prices down. The retailer can therefore operate with a mark-up, that is greater than his own variable selling costs k and still achieve the joint profit maximising outcome.

Formally, in the model of section 2, the retailer's problem in the first stage is to set a lump sum payment L^* and a percentage mark-up r^* so as to achieve the joint profit maximising price p^* (and output x^*) and then extract all the producer's surplus.

From (15) we know that the producer will set his own mark-up equal to

$$m = c/(\varepsilon - 1) \tag{15}$$

i.e. the wholesale price will equal

$$p_w = \varepsilon c/(\varepsilon - 1) > c. \tag{36}$$

Since the joint profit maximising price p^* is equal to

$$p^* = \varepsilon(c + k)/(\varepsilon - 1) \tag{10}$$

the retailer must set his percentage mark-up to equal

$$r^* = k/c. \tag{37}$$

The absolute value of the retailer's mark-up is thus

$$r^* = k/(\varepsilon - 1) > k. \tag{36}$$

The retailer can then set a fixed lump sum payment so as to extract all the producer's profits. This lump sum payment would equal

$$L^* = x^*c/(\varepsilon - 1). \tag{37}$$

The important point to note is that in this joint profit maximising contract *both the producer and the retailer* charge positive mark-ups, and in the case of the retailer this mark up is greater than his marginal selling cost k. Both market participants bear part of the risk and both have an incentive to see to it that the sales volume x* can in fact be achieved. This is a marked improvement over the contracts with absolute mark-ups where the one side bears all the risk, and the other side has serious moral hazard problems.

5. CONCLUSION

The purpose of this paper was to study the incentive effects of the commonly observed retailer policy of charging given percentage mark-ups. It was shown that the first obvious effect is to allow the retailer to squeeze the producer's mark-up and keep a greater share of aggregate profits for himself. However, the story does not end there. Since percentage mark-ups tend to render the demand function the producer faces more elastic, they also contribute to partially solving the double marginalization problem.

When the producer can extract all the retailer profits he may prefer to work with retailers pursuing a percentage mark-up strategy. In the absence of fixed transfer payments, they once again help in approaching the joint profit maximising solution. However, in this case it is the producer who gets most of the benefits.

Percentage mark-ups also affect the nature of the first best contracts, when lump-sum payments are possible. They permit a more even spreading of the risks between producer and retailer and my thus help alleviate moral hazard problems.

It may be interesting to compare the approach developed here with the more traditional literature on «rule of thumb» pricing.[6] The focus there is usually to examine whether the pricing rules used (such as the percentage mark-up rule analysed here) is compatible with profit maximisation or not. In all of those studies, no attempt is made to study the strategic effects of the pricing rule being applied. This could be a major oversight. It may of course be true that firms work with percentage mark-ups simply because they are easy to use. This paper has, however, illustrated that they may also be a powerful instrument

6. C.f. for example the survey by SILBERSTON (1970).

to exercise market power and can lead to more efficient outcomes than the simple absolute mark-ups that occupy such a central place in industrial organisation textbooks. Maybe these are the real reasons why one observes them so frequently in the real world.

REFERENCES

SCHERER, F. M. and D. ROSS (1990), Industrial market structure and economic performance, Boston: Houghton Mifflin.

SILBERSTON, A. (1970), Surveys of applied economics: Price behaviour of firms, *Economic Journal* 80, 511–582.

SPENGLER, J.J. (1950), Vertical integration and antitrust policy, *Journal of Political Economy* 58, 347–352.

TIROLE, J. (1988), The theory of industrial organisation, Cambridge (Mass.): MIT Press.

SUMMARY

A common assumption in the literature on the double marginalization problem is that the retailer can set his mark-up only in the second stage of the game after the producer has moved. To the extent that the sequence of moves is designed to reflect the relative bargaining power of the two parties it is just as plausible to let the retailer move first. Furthermore, retailers frequently calculate their selling prices by adding a percentage mark-up to their wholesale prices. This allows a retailer to obtain higher profits than with a standard linear pricing policy. This result holds both for the case of simultaneous moves and when the retailer moves first. Under such a «percentage mark-up» strategy equilibrium prices are lower and aggregate producer-plus-retailer profits are higher than under linear pricing. This last result has implications for situations, where there is intense competition among retailers. In the absence of franchise fees (which are quite rarely observed in practice), producers may prefer working with retailers using percentage mark-ups.

ZUSAMMENFASSUNG

In der Literatur zum Problem der «doppelten Marginalisierung» wird im Allgemeinen angenommen, dass der Händler seinen Aufschlag nur in der zweiten Stufe des Spiels nach dem Produzenten festsetzen kann. Da die Sequenz in mehrstufigen Spielen häufig auch die relative Verhandlungsmacht der Parteien widerspiegelt, ist die umgekehrte Reihenfolge ebenso plausibel. Wenn der Händler seine Preise festsetzt, indem er einen prozentualen Aufschlag auf die Grosshandelspreise nimmt, kann er höhere Gewinne erzielen als mit dem üblichen absoluten Aufschlag (mit dem in theoretischen Modellen so oft gearbeitet wird). Dies gilt sowohl in simultanen als auch in zweistufigen Spielen.

«Prozentuale Aufschläge» führen zu niedrigeren Konsumentenpreisen und höheren aggregierten Gewinnen. Dies hat auch Implikationen für Situationen, wo intensiver Wettbewerb zwischen den Händlern herrscht. Wenn Konzessionsgebühren nicht möglich sind, werden die Produzenten eventuell lieber mit Händlern arbeiten, die prozentuale Aufschläge nehmen.

RESUME

Dans la littérature sur le problème de la «double marginalisation» on suppose souvent, que le détaillant ne peut fixer sa marge que dans la deuxième étape du jeu, après le producteur. La séquence des mouvements dans les jeux à plusieurs étapes reflète souvent le pouvoir de négociation des parties. La séquence inverse est donc tout aussi plausible. Si le détaillant fixe son prix en ajoutant une marge proportionnelle au prix de gros il peut obtenir un bénéfice plus important qu'avec une marge absolue (qu'on emploie souvent dans les modèles théoriques). Cela vaut pour des jeux simultanés et pour des jeux séquentiels. Les marges proportionnelles conduisent à des prix finaux plus bas et des bénéfices agrégés plus importants. Cela a des implications pour des situations, où il y a une forte concurrence entre les détaillants. Dans l'absence de frais de franchise le producteur préférera éventuellement collaborer avec un détaillant que demande une marge proportionnelle.