Competitive Hold-Up: Monopoly Prices and the Response in Related Markets

August 12, 2017

Eric Rasmusen

Abstract

Hold-up creates an unappreciated cost of monopoly even if the held-up side of the market is perfectly competitive. If a monopolist cannot commit to a wholesale price in advance, competitive retailers with U-shaped cost curves will be reluctant to enter, knowing that the monopolist has incentive to raise the price and reduce their quasi-rents. Forward-looking retailers will earn zero profits in the long run, but their caution hurts the monopolist by shifting in the short-run market supply of retailer services. A similar problem occurs if the monopolist's product is sold directly to consumers but is a complement to a perfectly competitive good. Competitive hold-up arises from upstream opportunism, not downstream market power, and so is distinct from double marginalization and the two-monopoly complements externality.

Rasmusen: Dan R. and Catherine M. Dalton Professor, Department of Business Economics and Public Policy, Kelley School of Business, Indiana University. BU 438, 1309 E. 10th Street, Bloomington, Indiana, 47405-1701. (812) 855-9219. Fax: 812-855-3354. erasmuse@indiana.edu, http://www.rasmusen.org.

This paper: http://www.rasmusen.org/papers/holdup-rasmusen.pdf.

Keywords: Monopoly, holdup-costs, opportunism, competitive hold-up, quasi-rents, complements, double marginalization, retailers.
JEL Codes: D23, D42, D86, L14, L23.

I would like to thank Allan Collard-Wexler, Louis Kaplow, Qihong Liu, Michael Rauh and participants in Economics Job Market Rumors and meeting of the International Industrial Organization Society, Midwest Theory, the BEPP Brown Bag Lunch for their comments.

1. Introduction

This paper explores a monopoly pricing problem I will call "competitive hold-up" that arises because of the monopolist's temptation to choose his price so as to take advantage of perfectly competitive firms' fixed costs. The result is prices too high to maximize monopoly profits.

In the textbook model of U-shaped cost curves, the firm's fixed cost is sunk in the short run, and it earns revenue in excess of its variable costs. These excess revenues are "quasi-rents." If the market price falls unexpectedly the firm still stays in operation in the short run, with positive though lower-than-expected quasi-rents. In the long run some firms exit, the market price rises, and profits return to zero. In competitive holdup, the monopolist tries to take advantage of the short-run fixity of the number of competitive firms. He faces a tradeoff between a high price and a larger number of firms on the other side of the market. This tradeoff generates an optimal price. The problem is that once that optimal number of firms have entered, the monopolist's decision is different. He now will face just the tradeoff between a high price and higher sales to each firm. The number of firms being fixed, the monopolist will choose a higher price. This would generate negative profits for the competitive industry, so the number of firms entering will not be that optimal number, but a smaller number appropriate to the price they know the monopolist will want to charge in the short run for that size of industry. Competitive holdup is self-defeating, because it shrinks the competitive industry in the long run. Fewer firms will enter the competitive industry than if the monopolist could commit not to use the tactic.

This is much like the standard hold-up problem, but with a competitive industry on one side of the market. Hold-up costs are well understood in the context of contracting with one buyer and one seller. The buyer may refrain from undertaking value-increasing investments for fear that the seller will raise the price once the investment is a sunk cost. Williamson (1975) proposes this as a major force in explaining interactions in industrial organization, and Hart & Moore (1988) provides a formalization. Long-term contracts with various kinds of special clauses are one solution, explored in Noldeke, Georg & Schmidt (1995) (option clauses), Edlin & Reichelstein (1996) (damage clauses) and elswhere. The problem has been examined in empirical work as well, most famously in Joskow (1987) on the case of specific investments in delivering coal to power plants. In the present paper, the specific investment is a firm's fixed cost, the fixed cost in the textbook model of U-shaped cost curves. Since marginal cost is rising, not constant, the market is not completely shut down by opportunism and can operate even without the bilateral contracting which is the focus of the hold-up literature. Indeed, when firms are atomistic, the transactions cost of contracting makes it more impractical as a solution. Thus, we have another cost of monopoly besides the conventional allocative and rent-seeking losses.

We will look at two settings. The first will be the "retailer model", in which the monopolist sells his product through retailers who are perfectly competitive (sections 2 and 3). The second will be the "complements model", in which the monopolist sells directly to consumers but consumers also buy a complement good produced by a perfectly competitive industry (sections 4 and 5).

2. The Retailer Model

Let a monopolist produce a good at a constant marginal cost a in a market whose consumers demand quantity $Q_d(p)$ at retail price p. We will assume that $2Q_d' + (p-a)Q_d'' < 0$ so monopoly profit will be concave in price.[1] The monopolist sells quantity $q(t)$ at a wholesale price of w per unit to retailer t, The retailer resells to consumers at price p, incurring a marginal service cost of $c(q(t))$ to do so, where c is increasing. Each retailer must also incur a fixed cost of F. Retailers are infinitesimal, following Aumann (1964), Novshek & Sonnenschein

[1]This is a version of the Novshek Condition of Novshek (1985) and Gaudet & Salant (1991), which is commonly used in imperfect competition models to ensure that the monopoly pricing problem is convex.

(1987), and Peck (2001). The amount of retailers is n, so their output is $\int_0^n q(t)dt$ and the amount of fixed cost is nF. We will suppress the t argument, since in equilibrium all active retailers will choose the same output. Under these assumptions the competitive industry is a constant-cost industry; an individual firm's costs do not increase as the size of the industry grows, so long-run supply is perfectly elastic.

We will start with the retailer's decisions, and then move to the monopolist's. First, take the quantity n of retailers and the wholesale price w as given. The profits of an individual retailer are

$$\pi_{retailer} = pq - \left(F + \int_0^q c(x)dx + qw\right) \tag{1}$$

Maximizing by choice of output, the retailer's first order condition yields the equilibrium output q, at which price equals marginal cost:

$$p = c(q) + w, \tag{2}$$

Rearranging gives us the the short-run individual supply curve

$$q(p) = c^{-1}(p - w) \tag{3}$$

and the market supply curve

$$Q^s(p) = nc^{-1}(p - w). \tag{4}$$

The market must clear, so quantity demanded must equal quantity supplied. Since all n retailers are identical, their first order conditions and equilibrium outputs are the same and we can write:

$$Q_d(p) = nq, \tag{5}$$

or, using equation (2),

$$Q_d(c(q) + w) = nq. \tag{6}$$

The short-run equilibrium output of a firm, $q^*(n, w)$ (as distinguished from the firm's short-run supply curve $q(p)$) is therefore

$$q^*(w, n) = \frac{Q_d(c(q^*) + w)}{n}. \tag{7}$$

Lemma 1: *Individual retailers supply less if the number of retailers n or the wholesale price w is greater. The market equilibrium quantity, $Q^*(n,w)$, increases with n.*

$$\frac{\partial q}{\partial n} < 0, \quad \frac{\partial q}{\partial w} < 0, \quad \frac{\partial Q^*}{\partial n} > 0.$$

Proof. Differentiating equation (7) yields the individual retailer comparative statics we need for n and w. Differentiating with respect to n holding w constant yields

$$\frac{\partial q}{\partial n} = \frac{Q_d' c'(q) \frac{\partial q}{\partial n}}{n} - \frac{q}{n^2} \tag{8}$$

so

$$\frac{\partial q}{\partial n}\left(\frac{Q_d' c'(q)}{n} - 1\right) = \frac{q}{n^2} \tag{9}$$

and

$$\frac{\partial q}{\partial n} = \frac{q}{n^2}\left(\frac{n}{Q_d' c'(q) - n}\right) \tag{10}$$

Since $Q_d' < 0$, we can conclude that $\frac{\partial q}{\partial n} < 0$.

Differentiating equation (7) with respect to w holding n constant yields

$$\frac{\partial q}{\partial w} = \frac{Q_d' c'(q) \frac{\partial q}{d\partial}}{n} + \frac{Q_d'}{n} \tag{11}$$

so

$$\frac{\partial q}{\partial w}\left(\frac{Q_d' c'(q)}{n} - 1\right) = -\frac{Q_d'}{n} \tag{12}$$

and

$$\frac{\partial q}{\partial w} = -\frac{Q_d'}{Q_d' c'(q) - n} \tag{13}$$

Since $Q_d' < 0$, we can conclude that $\frac{\partial q}{\partial w} < 0$.

For the effect of n, note that retailer market supply, equation (4), tells us that as n increases, for a given price p it follows that market quantity supplied will increase too. The increase in quantity will cause p to fall since the demand function $Q_d(p)$ is downward sloping, but since both demand and supply functions are continuous, output will equilibrate at a level greater than the initial one, and $\frac{dQ^*}{dn} > 0$. ■

We can now consider the long-run equilibrium, in which the quantity n of firms is determined. Since there is free entry and perfect foresight, long-run retailer profits equal zero:

$$\pi_{retailer} = pq - \left(F + \int_0^q c(x)dx + qw \right) = 0 \tag{14}$$

We can solve for the retail price, which must equal average cost:

$$p = average\ cost = \frac{F}{q} + \frac{\int_0^q c(x)dx}{q} + w. \tag{15}$$

Since retailer profits are zero and retailers are identical, in equilibrium they will all choose the quantity that minimizes average cost, the minimum of the U-shaped average cost curve. That value, which we will call q^*, is the value that minimizes the average cost in equation (15) by solving the first order condition,

$$\frac{-F}{(q^*)^2} - \frac{\int_0^{q^*} c(x)dx}{(q^*)^2} + \frac{c(q^*)}{q^*} = 0. \tag{16}$$

Equation (2) says that $p = c(q) + w$, so since output per firm is fixed at q^*, independently of w, it follows that $\frac{dp}{dw} > 0$. Since market demand slopes down ($Q_d' < 0$) and since quantity supplied is nq^* in equilibrium it follows that $\frac{dn}{dw} < 0$, as stated in Lemma 2.

Lemma 2: *In the long run, the number of retailers, n, is decreasing in the wholesale price, w.*

$$\frac{dn}{dw} < 0.$$

When the wholesale price rises, the retail price must rise if the retailers are not to have negative profits, and since each retailer must produce at the cost-minimizing size, the number of retailers must fall.

The Upstream Monopolist

The upstream monopolist's profit is

$$\pi_{monopoly} = (w - a) \cdot n(w) \cdot q^*(n(w), w), \tag{17}$$

where the equilibrium output function $q^*(n(w), w)$ depends on the number of firms and the wholesale price (which generates the retail price to which the retailers respond directly). In the long run, sales per retailer are $q^*(n(w), w)$, determined entirely by w. In the short run, sales per retailer are $q^*(n, w)$, with n fixed by previous entry.

If the monopolist chooses w after retailers enter, he takes n as given and uses the short-run equilibrium sales function $q^*(n, w)$. Since the retail equilibrium price equals $p^* = c(q^*) + w$, monopoly profit (17) is concave in w as a result of our assumption that $Q_d(p)$ is concave. The first order condition is

$$\frac{d\pi_{monopoly}}{dw} = nq^*(n, w) + (w - a)n\frac{\partial q^*}{\partial w} = 0, \tag{18}$$

so

$$w = a - \frac{q^*(n, w)}{\frac{\partial q^*}{\partial w}}. \tag{19}$$

Thus, the monopolist's wholesale price equals his unit cost, a, plus (since $\frac{\partial q^*}{\partial w} < 0$) an amount depending on how much wholesale demand falls with the price, which in turn depends on the retail demand.

If the monopolist chooses w *before* retailers enter, he uses the long-run equilibrium sales function $q^*(n(w), w)$. Since now n is endogenous, we no longer know that monopoly profit is concave, but it is differentiable in w and the optimum will not be zero or infinity under our assumptions, so the first order condition is still a necessary, if not sufficient, condition and we can compare it to the short-run optimum. The first order condition is

$$\frac{d\pi}{dw} = nq^*(n, w) + (w - a)n\frac{\partial q^*}{\partial w} + (w - a)\frac{\partial}{\partial n}[nq^*(n, w)]\frac{dn}{dw} = 0, \tag{20}$$

so

$$w = a - \frac{q^*(n, w)}{\frac{\partial q^*}{\partial w} + \boxed{\frac{1}{n}\frac{\partial Q^*}{\partial n}\frac{dn}{dw}}}. \tag{21}$$

Lemma 1 tells us that $\frac{\partial q^*}{\partial w} < 0$ and $\frac{dQ}{dn} > 0$. Lemma 2 tells us that $\frac{dn}{dw} < 0$. Therefore, the boxed term in the denominator is negative, which makes the negative quantity in the denominator bigger in magnitude, so less is added to a to get the wholesale price. The monopolist who can commit to w therefore chooses a lower value than if he could not commit.

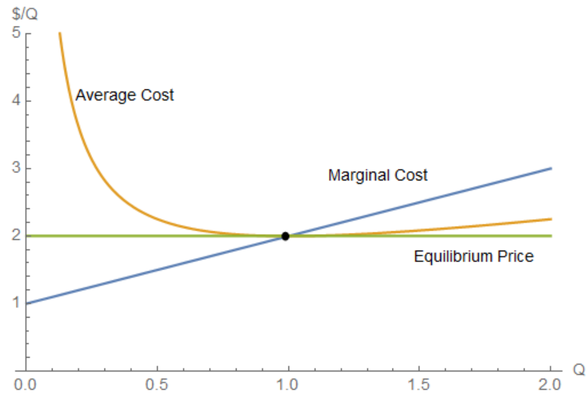
The monopolist who cannot precommit to a wholesale price will choose a higher price and sell less to retailers. Thus, the retailer price will be higher, and consumer surplus lower. Since the value of $n(w)$ will depend on the anticipated w anyway if retailers are rational, the monopolist's neglect of that term in the short-run profit maximization problem will reduce his profits. Thus, we have Proposition 1.

Proposition 1: *If the upstream monopolist cannot precommit to his wholesale price before retailers enter, his price will be higher, sales lower, and profits lower. As consumer surplus will also be lower, total surplus is lower than in a monopoly which can commit to wholesale prices.*

3. Example and Discussion

An example helps to show how the wholesale price, retailer output, and number of retailers interact. Let consumer demand be given by $Q(p) = 1000 - 200p$ and let the upstream monopolist produce at constant marginal cost $a = 1$ and sell at wholesale price w. Let there be a continuum of length n of competing retailers each selling q (if we may be loose technically in our phrasing) with a fixed cost of .5 and marginal cost of $c(q) + w$ where $c(q) = q$. A retailer's total cost is thus $.5q^2 + wq + .5$, with the U-shaped cost curve shown in Figure 1. The minimum average cost is $q = 1$.

Figure 1: A Retailer's Cost Curves



For the social optimum, start with the social marginal cost of serving consumers equalling the minimum average cost when the wholesale price is set at the wholesale marginal cost, so $w = a = 1$. Then, since $q = 1$, the average cost is $.5(1) + 1 + \frac{.5}{1} = 2$. The market quantity demanded at a price of $p = 2$ is 600. Dividing by q yields the quantity of retailers at the social optimum, $n = 600$. This will yield zero profits to the monopolist since $w = a$ and zero profits to the retailers since they are selling at average cost. Consumer surplus is 900.

To find the competitive equilibrium, begin with retailer behavior for a given wholesale price. A retailer's profit is

$$\pi_{retailer} = pq - \int_0^q c(x)dx - wq - F \tag{22}$$

Maximizing with respect to quantity yields price equals marginal cost:

$$p = c(q) + w \tag{23}$$

In equilibrium, market supply equals market demand, so

$$Q_s = nq = Q_d = 1000 - 200p \tag{24}$$

In the short run, n is fixed. With our cost function, price equalling marginal cost tells us that $p = q+w$, so the individual retailer short-run supply curve is

$$q(p, w) = p - w. \tag{25}$$

Solving equation (24) for p and substituting into (25) yields $q = 5 - n/200 - w$, which leaves us with an expression for the individual retailer's output as a function of n and w when p takes its resulting short-run equilibrium value:

$$q(n, w) = \frac{5 - w}{1 + .005n}. \tag{26}$$

In the long run, the amount of retailers n is determined by profits equalling zero. Since we have assumed $c(q) = q$, the price is $p = q + w$, which whe inserted into the retailer's profit function (22) yields the long-run equilibrium condition,

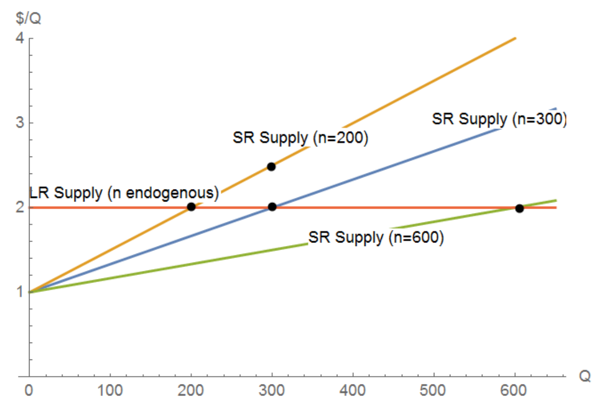
$$\pi_{retailer} = (q + w)q - .5q^2 - wq - F = 0. \tag{27}$$

The wq terms cancel, and we have assumed an entry cost of $F = .5$, so this solves to $q = 1$. Substituting $q = 1$ into the market equilibrium condition (24) gives us the equilibrium amount of retailers as a function of the wholesale price:

$$n^* = 800 - 200w. \tag{28}$$

As Figure 2 shows, the market supply curve swivels down as the number of retailers n increases, since to increase the quantity the market supplies requires moving less far up the individual retailer's marginal cost curves.

Figure 2: How Market Supply Changes with More Retailers



Assume first that the monopolist can commit to a wholesale price. He will substitute for n using the retailer long-run equilibrium condition, equation (28):

$$\begin{aligned} \pi_{monopolist} &= (w-a)qn \\ &= (w-1)(1)(800-200w) \end{aligned} \tag{29}$$

Solving for the optimal w yields $w = 2.5$, in which case $p = 3.5, n = 300, q = 1$ and $Q = 300$. The monopolist earns 450 and consumer surplus is 225.

Now suppose the monopolist does not choose w until after the retailers have made their entry decisions. He takes n as given and uses the retailers' short-run supply function (26) to form his expectation of q. Thus,

$$\begin{aligned} \pi_{monopolist} &= (w-a)nq(n,w) \\ &= (w-1)n(\tfrac{5-w}{1+.005n}) \\ &= \left(\tfrac{n}{1+.005n}\right)(5w-5-w^2+w) \end{aligned} \tag{30}$$

The first order condition for choice of w is

$$\frac{d\pi_{monopolist}}{dw} = \left(\frac{n}{1+.005n}\right)(6-2w) = 0 \tag{31}$$

Thus, the monopolist will choose a wholesale price of $w = 3$. The level of n will anticipate the monopolist's decision. We know that to achieve zero profit, a retailer must be operating at the minimum efficient scale of $q = 1$ with marginal cost of 1, so $p = w + 1 = 4$. The demand curve tells us that $Q_d = 1000 - 200p = 200$, which when $q = 1$ means that $n = 200$ also. The monopoly profit equals 400, less than the 450 with commitment, and consumer surplus equals 100, also less.

On the other hand, consider what happens if the monopolist is able to get away with lying to retailers, promising them the commitment wholesale price of $w = 2.5$ but actually being free to change. As we have seen, if they believe that $w = 2.5$ then amount $n = 300$ of retailers will enter, expecting zero profits. Now, however, the monopolist will rely on their short-run supply curve and use first-order condition (31), which yields $w = 3$. The retailers will use their short-run supply curve (26) and choose

$$q(n, w) = \frac{5 - w}{1 + .005n} = \frac{5 - 3}{1 + 1.5} = .8, \tag{32}$$

so output will be $Q = (300)(.8) = 240$ and $p = 5 - 2.4/2 = 3.8$. The monopoly's profit will be 480, and consumer surplus will be 144. Retailer profits will be negative, equalling

$$pQ - n(.5q^2 + wq + F) = 912 - 300(.5 * .8^2 + 3 * .8 + .5) = -54 \tag{33}$$

Note that in the no-commitment world, industry profit and consumer surplus are higher if the retailers are deceived than if they are not. Industry profit (monopoly profit minus retailer loss) comes to 426 instead of 400, while consumer surplus is 144 instead of 100. The behavior of forward-looking retailers leads to industry output below the level that maximizes industry profit, which itself is below the level that maximizes total surplus. If retailers are myopic, more enter, and that reduces the cost of producing market output levels closer to the industry optimum. Market output increases, leading to an overall gain in surplus— but a loss to retailers.

Table 1 collects the various outcomes. When the monopolist cannot commit to a wholesale price, he sets the wholesale price higher, fewer retailers enter, retail prices are higher, and monopoly profits fall from 450 to 400. If he faces myopic retailers who expect the commitment price of $w = 2.5$, he earn profits of 480. This is the only case in which retailers operate at any scale except the efficient one; they produce .8 instead of 1 because they shrink their output to reduce marginal cost in response to the unexpectedly high wholesale price.

TABLE 1:
EQUILIBRIUM VALUES IN THE NUMERICAL EXAMPLE

	Monopoly with commitment	Monopoly without commitment	Monopoly with deception	Social Optimum
Wholesale price, w	2.5	3	3	1
Retail price, p	3.5	4	3.8	2
Amount of retailers, n	300	200	240	600
Output per retailer, q	1	1	.8	1
Monopolist profit	450	400	480	0
Retailer profit	0	0	-54	0
Consumer surplus	225	100	144	900
Total surplus	675	500	570	900

For a given level of total output there is no inefficiency in production as the result of hold-up. Retailers operate at the efficient scale of q^* and the only problem is that there are too few of them. If the retailers were, instead, surprised by the high wholesale price, then too many would enter and they would end up each producing at too small a scale. Farsighted retailers foresee the high wholesale price, so just enough enter to earn zero profit at the minimum efficient scale. Inefficiency arises not from each retailer producing the wrong amount, but from the quantity of retailers being too small for the level of consumer demand. If the industry tried to sell the commitment or first-best output using the no-commitment number of retailers, the cost would be higher than if more retailers entered, because each retailer would have to expand his output beyond q^* up the marginal cost curve.

Comparison to double marginalization and secret quantity expansion

Competitive hold-up, like double marginalization features a monopolist who charges a price higher than the price which maximizes industry profits. The idea of double marginalization dates back to Spengler (1950) and in more modern form is described in Greenhut & Ohta (1979) and Janssen & Shelegia (2015). If an upstream manufacturer sells to downstream retailers who resells to consumers, they each add their own profit margin to the price they charge. When the retailer chooses a higher price, however, he ignores the fact that the resulting reduction in quantity sold hurts the manufacturer by reducing the quantity of input the retailer needs. Each of the two layers of sellers imposes a negative externality on the other by choosing a high profit margin that reduces the quantity ultimately sold to the consumer. Both would be better off with a contract under which the manufacturer reduced the wholesale price in return for the retailers reducing retail prices. Vertical integration also would work, since if the two firms merged, the externality would be internalized. The internal wholesale transfer price would be set to marginal cost and the monopoly retail price would be lower because of the lower wholesale price.

The problem— prices higher than even the seller desire—is the same for competitive hold-up, as are the clearest solutions— contracting over prices, and vertical integration. The optimal contract would fix the price in advance of retailer entry. Vertical integration would also work, because the merged firm would use the same minimum efficient scale quantity for each retail outlet and treat the resulting minimum average cost as the marginal cost of providing retailing services. The amount of retail outlets would then be chosen to maximize industry profits.

The difference is that double marginalization arises because both manufacturer and retailer have market power, while hold-up arises because of the U-shaped cost curve. If the retailers have no market power, the difference between the wholesale and retail prices is entirely due to the cost of retail services. There is no bargain to be made for both sides to lower their profit margins in order to increase quantity because the retailers have no profit margins to lower— they would have to price below minimum average cost. Hold-up, in contrast, is based on quasi-rents rather than rents and can occur either with or without retailer market power. The idea is that the manufacturer's price determines the amount of entry, which will be smaller if retailers anticipate the price to be higher. The model with perfectly competitive retailers gives the cleanest results and prevents hold-up from being confused with the double marginalization that would occur if retailers had market power. Since retailers will earn zero profits in equilibrium, any increase in the wholesale price will reduce the amount of retailers. If retailers had both U-shaped cost curves and market power, whether hold-up would reduce profits to below zero and reduce retailer quantity would depend on the initial level of retailer profits and other details of the model, while double marginalization would by itself mean that the price would be above the level chosen under optimal contracts or vertical integration.

A different solution to double marginalization, the two-part tariff, merely exacerbates holdup. It helps with double marginalization because the upstream monopolist can charge a unit price equal to his own marginal production cost to provide the retailers with the proper quantity incentive and a fixed price to extract their profits. Under competitive hold-up, the monopolist would choose the unit price to equal his own marginal production cost and the fixed price to seize all the retailer's quasi-rents. Foreseeing this, no retailer would enter. The possibility of using a two-part tariff would thus be purely bad for the monopolist. The problem in double marginalization is that the retailers are making monopoly profits at the expense of the manufacturer, so a solution is to transfer all the decisionmaking power to the manufacturer. An alternative is to give all the bargaining power to the retailers— the ability to make a take-it-or-leave it offer to buy the good at manufacturer marginal cost would solve double marginalization. In competitive holdup, the monopolist's bargaining leverage is the source of the problem; increasing it just squeezes more retailers out of the industry.

Competitive holdup is distinct from a different opportunism problem for a monopolist facing competitive retailers, what we might call "secret quantity expansion": after publicly agreeing to bilateral contracts with each retailer that would result in a certain total retail market quantity and price, the monopolist secretly sells more to some retailers. This would "cheat" the other retailers, who paid a wholesale price based on their belief that the retail price would be high because of the limited market quantity. Foreseeing this opportunism, the retailers will not accept a high price in the first place, dealing a blow to the monopolist's profits. The problem is reminiscient of the Coase Conjecture's competition between a monopolist present and future selves. This idea of secret quantity expansion can be found in Hart & Tirole (1990) in their discussion of vertical integration with two upstream and two downstream firms with identical goods, and is more clearly modelled as the central idea in the O'Brien & Shaffer (1992) model of a monopolist facing differentiated retailers. McAfee & Schwartz (1994) show that nondiscrimination clauses would not solve the problem, and Rey & Verge (2004) add close attention to the out-of-equilibrium beliefs of the retailers. Resale price maintenance does work as a solution, and Montez (2015) shows how product buybacks can serve a similar function.

Secret quantity expansion does not arise in the competitive holdup model; the opportunism is different. Here, the wholesale price is public and there is no bilateral contracting, but that is not the important difference. Rather, it is the assumption of upward sloping marginal cost for retailer services, which means that if the monopolist offers a greater quantity to a particular retailer, that retailer will have higher marginal cost and hence weaker demand than the others. Thus, the monopolist will not be tempted to offer secret discounts. If marginal costs were constant, on the other hand, and the monopolist had to use a per-unit price rather than a two-part tariff, that price would be above marginal cost, making retailers individually eager to be allowed to sell larger quantities and creating the temptation for secret quantity expansion.

4. The Complements Model

What we will next examine is the case of a monopolist who sells a good that is a complement of a good sold in a perfectly competitive industry. The firms will be unrelated on the supply side but linked by consumer preferences. Here, too, the monopolist will suffer from the temptation to raise prices to take advantage of the competitive industry's sunk cost.

Let there be two goods, one monopolized and one competitive. The monopolist produces the monopolized good at constant marginal cost a. Infinitesimal price-taking firms indexed by t produce quantity $y(t)$ of the competitive good by incurring a sunk fixed cost F and a nondecreasing marginal cost of $c(q(t))$. Firms are infinitesimal, so if the amount of firms is n their output is $\int_0^n y(t)dt$. As in the retailer model, we will suppress the t argument.

Let the monopolized good face demand $Q_d(w, r)$ at price w and the competitive firms face demand $Y_d(r, w)$ at price r. Note that w is now a price to consumers, not to another firm. The two goods are complements, so $\partial Q_d/\partial r < 0$ and $\partial Y_d/\partial w < 0$, but we will asume that quantity demanded is affected more by a good's own price, so $\partial Q_d/\partial w < \partial Q_d/\partial r < 0$ and $\partial Y_d/\partial r < \partial Y_d/\partial w < 0$. As in the retailer model, we will make the Novshek assumption so $2\partial Q_d/\partial w + (w - a)\partial^2 Q_d/\partial w^2 < 0$.

The Competitive Firm's Problem

First, take the quantity n of firms and the monopoly price w as given. The profits of an individual firm are

$$\pi_{firm} = ry - \left(F + \int_0^y c(x)dx \right) \tag{34}$$

Maximizing by choice of output, the retailer's first order condition yields the result that price equals marginal cost at the optimal output:

$$r = c(y), \tag{35}$$

which gives us the short-run individual supply curve,

$$y(r) = c^{-1}(r) \tag{36}$$

and a short-run market supply curve

$$Y_s(r, n) = nc^{-1}(r). \tag{37}$$

The market must clear, so quantity demanded must equal quantity supplied:

$$Y_d(r, w) = ny(r). \tag{38}$$

We can rewrite this as

$$Y_d(c(y), w) = ny(c(y)). \tag{39}$$

The short-run equilibrium output of a firm $y^*(n, w)$, as distinguished from the firm's short-run supply curve $y(r)$, is the y that solves the preceding equation:

$$y^*(w, n) = \frac{Y_d(w, c(y^*))}{n} \tag{40}$$

Lemma 3: *In the short run, if w rises then r falls, but at rate less than unity:*

$$-1 < \frac{dr}{dw} < 0$$

Proof. Totally differentiate the short run market supply equation (38) with respect to w, keeping n fixed and recognizing that the equilibrium value of r is a function $r(w)$:

$$\frac{\partial Y_d}{\partial w} + \frac{\partial Y_d}{\partial r}\frac{dr}{dw} = n\frac{dy}{dr}\frac{dr}{dw} \tag{41}$$

Rearranging, we get

$$\frac{dr}{dw} = \frac{\frac{\partial Y_d}{\partial w}}{n\frac{dy}{dr} - \frac{\partial Y_d}{\partial r}} \tag{42}$$

The numerator is $\frac{\partial Y_d}{\partial w}$, which is negative by assumption. The denominator is composed of two positive terms, since $\frac{dy}{dr}$ is positive and $\frac{\partial Y_d}{\partial r}$ is negative by assumption. Since, in addition, we assumed direct demand effects are bigger than cross-effects, we have $|\frac{\partial Y_d}{\partial w}| > |\frac{\partial Y_d}{\partial r}|$ and we can conclude that $-1 < \frac{dr}{dw} < 0$. ■

We can now consider the long-run equilibrium, in which the quantity n of firms is determined. Since there is free entry and perfect foresight, long-run competitive profits equal zero:

$$\pi_{firm} = ry - \left(F + \int_0^y c(x)dx\right) = 0 \tag{43}$$

We can solve this equation for the long-run price of the competitive good, which must equal the minimum average cost. The average cost is

$$Average\ cost = \frac{F}{y} + \frac{\int_0^y c(x)dx}{y} \tag{44}$$

Since retailer profits are zero and retailers are identical, in equilibrium they will all choose sales y to be the value that minimizes average cost. We will denote that minimum average cost by r^*. The value of y that minimizes verage cost solves the first order condition,

$$\frac{-F}{y^2} - \frac{\int_0^y c(x)dx}{y^2} + \frac{c(y)}{y} = 0. \tag{45}$$

Note that the price of the complementary monopolized good does not appear in (45); the long-run equilibrium scale does not depend on the demand side. If the monopolist chooses a higher price w, retailers end up exactly the same size in long-run equilibrium. Moreover, the long-run equilibrium price r^* is independent of w and of n. It is determined entirely on the supply side, by the minimum average cost, at the efficient scale. The effect of increasing the price of the monopolized good is just to reduce the number of competitive firms and industry output.

The Monopolist's Problem

The monopolist's profit is

$$\pi_{monopolist} = (w - a)Q_d(w, r(w)) \tag{46}$$

In the long run, r equals r^*, the average cost at the efficient scale, and so does not vary with w. In the short run, r falls as w increases, by Lemma 3.

It will be convenient here to start with the case of the monopolist who can commit to w. When the monopolist chooses w before retailers enter, he knows he will not affect the competitive good's price, which will equal r^*. His first order condition is

$$\frac{\partial \pi_{monopoly}}{\partial w} = Q_d(w, r) + (w - a)\frac{\partial Q_d(w, r)}{\partial w} = 0, \tag{47}$$

so

$$w = a - \frac{Q_d(w, r}{\frac{\partial Q_d(w,r)}{\partial w}} \tag{48}$$

Since $\frac{\partial Q_d(w,r(w))}{\partial w} < 0$ from the assumption that the demand curve sloped down, this means that the monopolist's wholesale price equals his cost, a, plus an amount depending on the price sensitivity of demand.

Without commitment, in the short-run when the monopolist chooses w after retailers enter his choice will affect r. His first order condition is

$$\frac{d\pi_{monopoly}}{dw} = Q_d(w, r(w)) + (w - a)\frac{\partial Q_d(w, r(w))}{\partial w} + (w - a)\frac{\partial Q_d(w, r(w))}{\partial r}\frac{dr}{dw} = 0, \tag{49}$$

so

$$w = a - \frac{Q_d(w, r(w))}{\frac{\partial Q_d(w,r(w))}{\partial w} + \boxed{\frac{\partial Q_d(w,r(w))}{\partial r}\frac{dr}{dw}}} \tag{50}$$

This differs from the wholesale price with commitment via the boxed term. That term is positive because the fact that the two goods are complements tells us that $\frac{\partial Q_d(w,r(w))}{\partial r} < 0$ and Lemma 3 tells us that $\frac{dr}{dw} < 0$. The first term in the denominator is negative because the demand curve slopes down: $\frac{\partial Q_d(w,r(w))}{\partial w} < 0$. The first term is bigger in magnitude than the second because we assumed that cross-price effects are smaller than own-price effects ($|\frac{\partial Q_d(w,r(w))}{\partial r}| < |\frac{\partial Q_d(w,r(w))}{\partial w}|$) and Lemma 3 tells us that $|\frac{dr}{dw}| < 1$. As a result, the effect of the boxed term is to make the denominator smaller in magnitude, but still negative. The equilibrium price without commitment will therefore be higher than with commitment.

If the competitive firms are rational, they will foresee that w will be higher in the no-commitment case, so demand Y^d will be weaker and n must be smaller for each firm to sell y^* at r^* and earn zero profits. Therefore, in the equilibrium without commitment, y and r end up the same as when there is commitment. But we have seen that when r equals r^*, the price that solves the monopolist's optimization problem is lower than our no-commitment monopoly price, so the no-commitment monopoly must be earning lower profit.

In the long run, the monopolist cannot affect the equilibrium price of the competitive good. If he can commit to a price for his own good, then he will choose a value low enough to encourage the right amount of entry into the competitive-good market. If he cannot commit, then fewer competitive firms will enter. He will find it optimal to react to that with a higher monopoly price, but though that will discourage sales of the competitive good, the discouragement will have been factored into the entry decision and so n will be just right to allow for zero profits in the competitive-good industry. Thus, we have Proposition 2.

Proposition 2: *If the monopolist cannot precommit to his output before competitive firms enter the market for a complementary good, his output and profit will be lower. As consumer surplus will also be lower, total surplus is lower than in a vertically integrated monopoly or a monopoly that could commit to future output.*

One way to understand Proposition 2 is to think of the monopolist as an innovator who starts production knowing that he will stimulate a secondary market for a different, complementary good next period. Firms thinking of entering the secondary market will have to think about how much the monopolist will produce then. They know that he will produce more once the second market opens up. Once the second period arrives, though, the monopolist will be conferring a positive externality on the firms in the secondary market. Thus, he will not produce enough to maximize social surplus. Foreseeing this, fewer firms will enter the secondary market. But since the monopolist gets all the surplus in the end anyway, he is worse off because he neglects the externality.

Another way is to think of the benefits to the monopolist from successful deception. Suppose the monopolist promises to produce a certain amount once the secondary market opens up. After the market does open up, the monopolist will want to break his promise. He has already gotten the entry he wanted to stimulate so his own good would sell more, and the entrants will not leave (in the short run) even if their revenues does not cover their entry costs. So the monopolist will produce more than if the secondary market did not exist, but not enough to make profits positive there. Foreseeing this, fewer firms would enter the competitive market.

5. An Example for the Complements Model

Let consumer demand be linear in the markets for the monopolized and competitive good:

$$Q_d(w, r) = 2100 - 200w - 180r \tag{51}$$

$$Y_d(r, w) = 1400 - 200r - 150w \tag{52}$$

Assume that the monopolist's marginal cost is $a = 1$ and the competitive firms have fixed cost of $F = .5$, and marginal cost $c(y) = y$, so a retailer's total cost is $.5 + .5y^2$. In the competitive industry, free entry makes profits equal zero, so all firms produce at exactly the minimum average cost in long-run equilibrium. This is found by minimizing $\frac{.5}{q} + .5q$, and yields marginal cost of $r^* = 1$ and output of $q = 1$ for each firm.

The monopolist's profit is

$$Profit(monopolist) = Q(w, r(w))(w-a) = (2100-200w-180r(w))(w-1). \tag{53}$$

When the monopolist can commit, he will realize that $r = 1$ in the long run and maximize

$$Profit(monopolist, commitment) = Q(w, 1)(w-a) = (2100-200w-180)(w-1) \tag{54}$$

This is maximized at $w = 5.3$. In that case, $n = Y_d = 1400 - 200r - 150w = 405$. Profit is 3,698. The price in the competitive market is $r = 1$.

Now consider the monopolist without commitment. Since the marginal cost is $c(y) = y$, the individual supply curve when price equals marginal cost is $y_s = r$, and the market supply is $Y_s = nr$. Thus, we have $Y_d = 1,400 - 200r - 150w = Y_s = nr$, so $r(w) =$

$(1,400 - 150w)/(200+n)$. In that case, the monopolist's profit function is

$$\begin{aligned} Profit(monopolist, no-commitment) &= Q(w, r(w))(w-a) \\ &= (1,400 - 200w - 180r(w))(w-1) \\ &= (1,400 - 200w - 180\tfrac{1,400-150w}{200+n})(w-1), \end{aligned} \tag{55}$$

which is maximized at

$$w(n) = (9,050 + 115n)/(20(65+n)) \tag{56}$$

We found that $n = 405$ in the commitment case. If the competitive market is myopic, so $n = 405$, then $w \approx 5.92$. The price in the competitive market is $r(405, 5.6) \approx .85$. Profit is 3,757.

A price of $r = .85$ is not a long-run equilibrium, however, because the competitive firms earn negative profits at that price. Rather, the no-commitment equilibrium is where

$$r(n) = \frac{1,400 - 150w(n)}{200+n} = 1, \tag{57}$$

which solves to $n \approx 305$ and $w \approx 5.96$. This is a profit of 3,610.

Thus, if the competitive firms think the monopolist will not keep his promise to set $w = 5.3$ and fear he would raise it to 5.92 to get them to reduce their price from 1 to .85, fewer of them enter. Since fewer enter, the monopolist ends up charging 5.96, even more than the 5.92 that would result from being able to fool a myopic competitive market.

Discussion

Just as the retailer model brings to mind double marginalization, so the complements model brings to mind the overpricing of complements produced by two monopolies. Each monopolist marks his price up above marginal cost. When one of them increases his mark-up, he captures the entire gain but inflicts a negative externality on the other monopolist. The two firms would both have higher profits if they simultaneously reduced their prices. This idea goes back to Cournot (1838). More recent analyses can be found at Economides & Salop (1992), Feinberg & Kamien (2001), Dari-Mattiacci & Parisi (2006), and Spulber (2016). As with competitive hold-up, the problem is high prices, and either a contract to reduce both prices or an internalizing merger would solve the problem.

The conventional complements problem, however, crucially depends on both markets being monopolized. If one market is perfectly competitive, the price there equals minimum average cost, which will not change as a result of the monopoly's raising its price. The negative externality is still there, but its only effect is to reduce the quantity sold in the complement industry, not the price. After the monopolist raises his price, consumer demand curve for the complement will shift in, with enough firms exiting the industry that the remaining ones are just sufficient to supply the new demand curve at the same minimum average cost price as before. Since the net effect is to leave the price unchanged, there is not the same feedback as if the complement market were monopolized and responded by raising its own price. There is only one policy tool, the monopolist's price, so he has complete control.

This is similar to why double marginalization was not present in the retailer model. Indeed, double marginalization is like a complements problem where one good— retail services— is a perfect complement for the other. In the competitive hold-up model, in contrast, one complement is sold at minimum average cost in equilibrium no matter what the monopolist does. The problem is one-sided.

6. Concluding Remarks

The key to competitive hold-up is that the price charged for the competitive good will equal the minimum average cost for that good regardless of the monopolist's price, so long as the monopolist's choice of price is foreseen. What the monopolist's choice of price does is determine how many firms enter and sell the competitive good.

This takes different forms in the retailer model and in the complements model. In the retailer model, the upstream monopolist's attempt at an opportunistic wholesale price means that to obtain the quasi-rents the retailers need, few enough of them will enter that industry retailer sales will be smaller and the retail price will be enough to cover both the high wholesale price and their services' cost. The smaller retail sales, however, are worse for the monopolist than if he were able to commit to a low wholesale price. In the complements model, the monopolist's high price for his good reduces the number of firms which can survive in the market for the competitive complement by weakening its demand. The monopolist would do better with a lower price and a greater number of competitive firms, which then as an industry would find it cheaper to supply a greater quantity of the complement to his own good.

Though the retailers and complement sellers in this paper have been perfectly competitive, competitive holdup would arise in the same way if they were monopolistically competitive, with market power but free entry. Such an industry would also be prone to double marginalization, but the price-raising effect of competitive hold-up would be added on top. And hold-up could even affect the number of firms when the other side of the market is oligopolistic rather than atomistic. The seizure of quasi-rents would not necessarily drive a firm's profits negative, since it would start with positive profits, but in some cases it would, and the number of retailers or complement sellers would decline.

This model helps us to understand the hold-up problem by showing that the relationship-specific investment can be the fixed cost of firms. Moreover, though one side of the market must have market power, the other can be atomistic and perfectly competitive. The inability to commit to a price in advance combined with sunk costs creates inefficiency just as it does in standard hold-up models, but the inefficiency is channelled through the number of firms. A competitive firm produces the same amount and adds the same mark-up for its costs whether the monopolist can commit or not; all that changes is the number of firms.

Is competitive hold-up a useful idea to apply to actual markets, as opposed to just understanding hold-up better? In the real world, there are few or no highly competitive industries that sell the products of one monopolist. The closest one might think of are franchises, which do use contracts to solve the hold-up problem. It is easier to think of examples of competitive complements where a firm's existence depends on a monopoly's supply— cases that fit only one brand of telephone, or after-market car parts, for example. Most companies in the real world, however, sell a variety of products. Competitive hold-up will still be present at the margin of a firm's decision of whether to stay in the market, even if the hold-up only raises one price in one hundred. Indeed, there will be a general tendency for the prices of the monopolized goods to take the number of firms retailing them or selling complements as unaffected by their prices, a tendency which one might otherwise (or in parallel) ascribe to bounded rationality. The firms with market power will look only to short-run elasticities of demand because they cannot commit to prices in the long run.

Thus, we can add hold-up to triangle losses, double marginalization, rent-seeking, aggravated agency costs (see, e.g., Farrell (2001) and other items on the list of the costs of market power. Hold-up also belong to the sub-list of inefficiencies arising from market power that hurt even the monopolist. Not only does it generate higher prices, but the higher prices fail to help the monopolist and to provide, for example, profits to incentivize the innovation that generates much of the market power in the economy.

References

Aumann, Robert J. (1964) "Markets with a Continuum of Traders," *Econometrica*, 32: 39–50.

Cournot, Augustin (1838) *Researches into the Mathematical Principles of the Theory of Wealth*, originally published in French as *Recherches sur les principes mathmatiques de la thorie des richesses.*

Dari-Mattiacci, Giuseppe & Francesco Parisi (2006) "Substituting Complements," *Journal of Competition Law and Economics*, 2: 333–347.

Economides Nicholas & Steven C. Salop (1992) "Competition and Integration among Complements, and Network Market Structure," *The Journal of Industrial Economics*, 40: 105-123.

Edlin, Aaron S. & Stefan Reichelstein (1996) "Holdups, Standard Breach Remedies, and Optimal Investment," *The American Economic Review*, 86: 478–501.

Farrell, Joseph (2001) "Monopoly Slack and Competitive Rigor," in *Readings in Games and Information*, Eric Rasmusen, ed., Oxford: Blackwell (2001).

Feinberg, Yossi & Morton I. Kamien (2001) "Highway Robbery: Complementary Monopoly and the Hold-up Problem," *International Journal of Industrial Organization*, 19: 1603–1621.

Gaudet, Gerard & Stephen W. Salant (1991) "Uniqueness of Cournot Equilibrium: New Results from Old Methods," *The Review of Economic Studies*, 58: 399–404.

Greenhut, Melvin L. & Hiromichi Ohta (1979) "Vertical Integration of Successive Oligopolists," *The American Economic Review*, 69: 137–141.

Hart, Oliver D. & Moore, John D. (1988) "Incomplete Contracts and Renegotiation," *Econometrica*, 56: 755–785.

Hart, Oliver & Jean Tirole (1990) "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity (Microeconomics),* 205–285.

Hunold, Matthias & Konrad Stahl (2016) "Passive Vertical Integration and Strategic Delegation," *The RAND Journal of Economics*, 47: 891–913.

Janssen, Maarten & Sandro Shelegia (2015) "Consumer Search and Double Marginalization," *The American Economic Review,* 105: 1683–1710.

Joskow, Paul L. (1987) "Contract Duration and Relationship-Specific Investments Empirical Evidence from Coal Markets," *The American Economic Review*, 77: 168–185.

Lafontaine, Francine & Margaret Slade (2007) "Vertical Integration and Firm Boundaries: The Evidence," *The Journal of Economic Literature*, 45: 629–685.

McAfee, R. Preston & Marius Schwartz (1994) "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity," *The American Economic Review*, 84: 210–230.

Montez, Joao (2015) "Controlling Opportunism in Vertical Contracting When Production Precedes Sales," *The RAND Journal of Economics*, 46: 650–670.

Noldeke, Georg & Schmidt, Klaus M. (1995) "Option Contracts and Renegotiation: A Solution to the Holdup Problem," *The RAND Journal of Economics*, 26: 163–179.

Novshek, William (1985) "On the Existence of Cournot Equilibrium," *The Review of Economic Studies,* 52: 85–98 (1985).

Novshek, William & Hugo Sonnenschein (1987) "General Equilibrium with Free Entry: A Synthetic Approach to the Theory of Perfect Competition," *Journal of Economic Literature*, 25: 1281–1306.

O'Brien, Daniel P. & Greg Shaffer (1992) "Vertical Control with Bilateral Contracts," *The RAND Journal of Economics*, 23: 299–308.

Peck, Richard M. (2001) "Infinitesimal Firms and Increasing Cost Industries," *The Journal of Economic Education*, 32: 41–52.

Rey, Patrick, Paul Seabright & Jean Tirole (2001) "The Activities of a Monopoly Firm in Adjacent Competitive Markets: Economic Consequences and Implications for Competition Policy," working paper, Institut d'Economie Industrielle, Universite de Toulouse-1 (September 21, 2001), http://idei.fr/sites/default/files/medias/doc/by/seabright/activities.pdf.

Rey, Patrick & Thibaud Verge (2004) Bilateral Control with Vertical Contracts, *The RAND Journal of Economics*, 35: 728–746.

Spengler, Joseph J. (1950) "Vertical Integration and Antitrust Policy," *The Journal of Political Economy*, 58: 347-352.

Spulber, Daniel F. (2016) "Complementary Monopolies and Bargaining," working paper, Northwestern University (June 8, 2016).

Williamson, Oliver E. (1975) *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press (1975).